HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human signal peptide-containing proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative disorders including cancer; inflammation; and cardiovascular, neurological, reproductive, and developmental disorders.

BACKGROUND OF THE INVENTION

Protein transport is essential for cellular function. Transport of a protein may be mediated by a signal peptide located at the amino terminus of the protein itself. The signal peptide is comprised of about ten to twenty hydrophobic amino acids which target the nascent protein from the ribosome to a particular membrane bound compartment such as the endoplasmic reticulum (ER). Proteins targeted to the ER may either proceed through the secretory pathway or remain in any of the secretory organelles such as the ER, Golgi apparatus, or lysosomes. Proteins that transit through the secretory pathway are either secreted into the extracellular space or retained in the plasma membrane. Secreted proteins are often synthesized as inactive precursors that are activated by post-translational processing events during transit through the secretory pathway. Such events include glycosylation, phosphorylation, proteolysis, and removal of the signal peptide by a signal peptidase. Other events that may occur during protein transport include chaperone-

a receptor or pore complex. Examples of secreted proteins with amino terminal signal peptides are discussed below and include receptors, extracellular matrix molecules, cytokines, hormones, growth and differentiation factors, neuropeptides, vasomediators, phosphokinases, phosphatases, phospholipases, phosphodiesterases, G and Ras-related proteins, ion channels, transporters/pumps, proteases, and transcription factors. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland Publishing, New York, NY, pp. 557-560, 582-592.)

G-protein coupled receptors (GPCRs) comprise a superfamily of integral membrane proteins which transduce extracellular signals. GPCRs include receptors for biogenic amines such as dopamine, epinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin; for lipid mediators of inflammation such as prostaglandins, platelet activating factor, and leukotrienes; for peptide hormones such as calcitonin, C5a anaphylatoxin, follicle stimulating hormone, gonadotropin releasing hormone, neurokinin, oxytocin, and thrombin; and for sensory signal mediators such as retinal photopigments and olfactory stimulatory molecules. The structure of these highly conserved receptors consists of seven hydrophobic transmembrane regions, cysteine disulfide bridges between the second and third extracellular loops, an extracellular N-terminus, and a cytoplasmic C-terminus. The N-terminus interacts with ligands, the disulfide bridges interact with agonists and antagonists, and the large third intracellular loop interacts with G proteins to activate second messengers such as cyclic AMP, phospholipase C, inositol triphosphate, or ion channels. (Reviewed in Watson, S. and Arkinstall, S. (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego, CA, pp. 2-6; and Bolander, F.F. (1994) Molecular Endocrinology, Academic Press, San Diego, CA, pp. 162-176.)

Other types of receptors include cell surface antigens identified on leukocytic cells of the immune system. These antigens have been identified using systematic, monoclonal antibody (mAb)-based "shot gun" techniques. These techniques have resulted in the production of hundreds of mAbs directed against unknown cell surface leukocytic antigens. These antigens have been grouped into "clusters of differentiation" based on common immunocytochemical localization patterns in various differentiated and

a single cell surface protein and are assigned a "CD" number. Some of the genes encoding proteins identified by CD antigens have been isolated and characterized as both transmembrane proteins and cell surface proteins anchored to the plasma membrane via covalent attachment to fatty acid-containing glycolipids such as glycosylphosphatidylinositol (GPI). (Reviewed in Barclay, A. N. et al. (1993) The Leucocyte Antigen Facts Book, Academic Press, San Diego, CA, pp. 144-145; Noel, L. S. et al. (1998) J. Biol. Chem. 273:3878-3883.)

Tetraspanins are a superfamily of membrane proteins which facilitate the formation

and stability of cell-surface signaling complexes containing lineage-specific proteins, integrins, and other tetraspanins. They are involved in cell activation, proliferation (including cancer), differentiation, adhesion, and motility. These proteins cross the membrane four times, have conserved intracellular – and C-termini and an extracellular, non-conserved hydrophilic domain. Tetraspanins include, e.g., platelet and endothelial cell membrane proteins, leukocyte surface proteins, tissue specific and tumorous antigens, and the retinitis pigmentosa-associated gene peripherin. (Maecker, H.T. et al. (1997) FASEB J. 11:428-442.)

Matrix proteins (MPs) are transmembrane and extracellular proteins which function in formation, growth, remodeling, and maintenance of tissues and as important mediators and regulators of the inflammatory response. The expression and balance of MPs may be perturbed by biochemical changes that result from congenital, epigenetic, or infectious diseases. In addition, MPs affect leukocyte migration, proliferation, differentiation, and activation in the immune response. MPs are frequently characterized by the presence of one or more domains which may include collagen-like domains, EGF-like domains, immunoglobulin-like domains, and fibronectin-like domains. In addition, some MPs are heavily glycosylated. MPs include extracellular proteins such as fibronectin, collagen, and galectin and cell adhesion receptors such as cell adhesion molecules (CAMs), cadherins, and integrins. (Reviewed in Ayad, S. et al. (1994) The Extracellular Matrix Facts Book, Academic Press, San Diego, CA, pp. 2-16; Ruoslahti, E. (1997) Kidney Int. 51:1413-1417; Sjaastad, M.D. and Nelson, W.J. (1997) BioEssays 19:47-55.)

Lectins are proteins characterized by their ability to bind carbohydrates on cell

(Kishore, U. et al. (1997) Matrix Biol. 15:583-592.) Certain cytokines and membrane-spanning proteins have CRDs which may enhance interactions with extracellular or intracellular ligands, with proteins in secretory pathways, or with molecules in signal transduction pathways. The lipocalin superfamily constitutes a phylogenetically conserved group of more than forty proteins that function by binding to and transporting a variety of physiologically important ligands. (Tanaka, T. et al. (1997) J. Biol. Chem. 272:15789-15795; and van't Hof, W. et al. (1997) J. Biol. Chem. 272:1837-1841.) Selectins are a family of calcium ion-dependent lectins expressed on inflamed vascular

endothelium and the surface of some leukocytes. (Rossiter, H. et al. (1997) Mol. Med. Today 3:214-222.)

Protein kinases regulate many different cell proliferation, differentiation, and signaling processes by adding phosphate groups to proteins. Reversible protein phosphorylation is a key strategy for controlling protein functional activity in eukaryotic cells. The high energy phosphate which drives this activation is generally transferred from adenosine triphosphate molecules (ATP) to a particular protein by protein kinases and removed from that protein by protein phosphatases. Phosphorylation occurs in response to extracellular signals, cell cycle checkpoints, and environmental or nutritional stresses. Protein kinases may be roughly divided into two groups; protein tyrosine kinases (PTKs) which phosphorylate tyrosine residues, and serine/threonine kinases (STKs) which phosphorylate serine or threonine residues. A few protein kinases have dual specificity. A majority of kinases contain a similar 250-300 amino acid catalytic domain. (Hardie, G. and Hanks, S. (1995) The Protein Kinase Facts Book, Vol I, pp. 7-47, Academic Press, San Diego, CA.)

Protein phosphatases remove phosphate groups from molecules previously modified by protein kinases thus participating in cell signaling, proliferation, differentiation, contacts, and oncogenesis. Protein phosphorylation is a key strategy used to control protein functional activity in eukaryotic cells. The high energy phosphate is transferred from ATP to a protein by protein kinases and removed by protein phosphatases. There appear to be three, evolutionarily-distinct protein phosphatase gene families: protein phosphatases (PPs); protein tyrosine phosphatases (PTPs); and acid/alkaline phosphatases (APs). PPs dephosphorylate phosphoserine/threonine residues

PTPs reverse the effects of protein tyrosine kinases and therefore play a significant role in cell cycle and cell signaling processes. Although APs dephosphorylate substrates in vitro, their role in vivo is not well known. (Charbonneau, H. and Tonks, N.K. (1992) Annu. Rev. Cell Biol. 8:463-493.)

Cyclic nucleotides (cAMP and cGMP) function as intracellular second messengers to transduce a variety of extracellular signals, including hormones, light and neurotransmitters. Cyclic nucleotide phosphodiesterases (PDEs) degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular

concentrations of cyclic nucleotides and their effects on signal transduction. At least seven families of mammalian PDEs have been identified based on substrate specificity and affinity, sensitivity to cofactors and sensitivity to inhibitory drugs. (Beavo, J.A. (1995) Physiological Reviews 75: 725-748.)

Phospholipases (PLs) are enzymes that catalyze the removal of fatty acid residues from phosphoglycerides. PLs play an important role in transmembrane signal transduction and are named according to the specific ester bond in phosphoglycerides that is hydrolyzed, i.e., A_1, A_2, C or D. PLA_2 cleaves the ester bond at position 2 of the glycerol moiety of membrane phospholipids giving rise to arachidonic acid. Arachidonic acid is the common precursor to four major classes of eicosanoids, namely prostaglandins, prostacyclins, thromboxanes and leukotrienes. Eicosanoids are signaling molecules involved in the contraction of smooth muscle, platelet aggregation, and pain and inflammatory responses. (Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland Publishing, Inc., New York, NY, pp. 85, 211, 239-240, 642-645.)

The nucleotide cyclases, i.e., adenylate and guanylate cyclase, catalyze the synthesis of the cyclic nucleotides, cAMP and cGMP, from ATP and GTP, respectively. They act in concert with phosphodiesterases, which degrade cAMP and cGMP, to regulate the cellular levels of these molecules and their functions. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals, e.g., hormones, and light and neurotransmitters. (Stryer, L. (1988) Biochemistry W.H. Freeman and Co., New York, pp. 975-980, 1029-1035.)

Cytokines are produced in response to cell perturbation. Some cytokines are produced as precursor forms, and some form multimers in order to become active. They

and the members of the group interact with one another and other molecules to produce an overall biological response. Interleukins, neurotrophins, growth factors, interferons, and chemokines are all families of cytokines which work in conjunction with cellular receptors to regulate cell proliferation and differentiation and to affect such activities as leukocyte migration and function, hematopoietic cell proliferation, temperature regulation, acute response to infections, tissue remodeling, apoptosis, and cell survival. Studies using antibodies or other drugs that modify the activity of a particular cytokine are used to elucidate the roles of individual cytokines in pathology and physiology.

Chemokines, in particular, are small chemoattractant cytokines involved in inflammation, leukocyte proliferation and migration, angiogenesis and angiostasis, regulation of hematopoiesis, HIV infectivity, and stimulation of cytokine secretion. Chemokines generally contain 70-100 amino acids and are subdivided into four subfamilies based on the presence of conserved cysteine-based motifs. (Callard, R. and Gearing, A. (1994) The Cytokine Facts Book, Academic Press, New York, NY, pp. 181-190, 210-213, 223-227.)

Growth and differentiation factors are secreted proteins which function in intercellular communication. Some factors require oligomerization or association with MPs for activity. Complex interactions among these factors and their receptors trigger intracellular signal transduction pathways that stimulate or inhibit cell division, cell differentiation, cell signaling, and cell motility. Most growth and differentiation factors act on cells in their local environment (paracrine signaling). There are three broad classes of growth and differentiation factors. The first class includes the large polypeptide growth factors such as epidermal growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, and platelet-derived growth factor. The second class includes the hematopoietic growth factors such as the colony stimulating factors (CSFs). Hematopoietic growth factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. The third class includes small peptide factors such as bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, and bradykinin which function as hormones to regulate cellular functions other than proliferation.

cells in vitro and in tumor progression in vivo. Inappropriate expression of growth factors by tumor cells may contribute to vascularization and metastasis of melanotic tumors. During hematopoiesis, growth factor misregulation can result in anemias, leukemias, and lymphomas. Certain growth factors such as interferon are cytotoxic to tumor cells both in vivo and in vitro. Moreover, some growth factors and growth factor receptors are related both structurally and functionally to oncoproteins. In addition, growth factors affect transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Reviewed in Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI, pp. 1-9.)

Proteolytic enzymes or proteases either activate or deactivate proteins by hydrolyzing peptide bonds. Proteases are found in the cytosol, in membrane-bound compartments, and in the extracellular space. The major families are the zinc, serine, cysteine, thiol, and carboxyl proteases.

Zinc proteases, e.g., carboxypeptidase A, have a zinc ion bound to the active site. These proteases recognize C-terminal residues that contain an aromatic or bulky aliphatic side chain, and hydrolyze the peptide bond adjacent to the C-terminal residues. Serine proteases have an active site serine residue and include digestive enzymes, e.g., trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of extracellular matrix (ECM) molecules. Cysteine proteases (e.g. cathepsin) are produced by monocytes, macrophages and other immune cells, and are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Overproduction of these enzymes can cause the tissue destruction associated with rheumatoid arthritis and asthma. Thiol proteases, e.g., papain, contain an active site cysteine and are widely distributed within tissues. Carboxyl proteases, e.g., pepsin, are active only under acidic conditions (pH 2 to 3).

Guanosine triphosphate-binding proteins (G proteins) can be grouped into two major classes: heterotrimeric G proteins and small G proteins. Heterotrimeric G proteins interact with GPCRs that respond to hormones, growth factors, neuromodulators, or other signaling molecules. The interaction between GPCR and G protein allows the G protein to exchange GTP for guanosine diphosphate (GDP). This exchange activates the G protein, allowing it to dissociate from the receptor and interact with the its cognate second or ion channels. The hydrolysis of GTP to GDP by the G protein acts as an on-off switch, terminating the action of the G protein and preparing it to interact with another receptor molecule, thus beginning another round of signal transduction.

The small G proteins consist of single 21-30 kDa polypeptides. They can be classified into five subfamilies: Ras, Rho, Ran, Rab, and ADP-ribosylation factor. These proteins regulate cell growth, cell cycle control, protein secretion, and intracellular vesicle interaction. In particular, the Ras proteins are essential in transducing signals from receptor tyrosine kinases to serine/threonine kinases which control cell growth and

differentiation. Mutant Ras proteins, which bind but can not hydrolyze GTP, are permanently activated and cause continuous cell proliferation or cancer. All five subfamilies share common structural features and four conserved motifs. Most of the membrane-bound G proteins require a carboxy terminal isoprenyl group (CAAX), added posttranslationally, for membrane association and biological activity. The G proteins also have a variable effector region, located between motifs I and II, which is characterized as the interaction site for guanine nucleotide exchange factors or GTPase-activating proteins.

Eukaryotic cells are bound by a membrane and subdivided into membrane-bound compartments. Membranes are impermeable to many ions and polar molecules, therefore transport of these molecules is mediated by ion channels, ion pumps, transport proteins, or pumps. Symporters and antiporters regulate cytosolic pH by transporting ions and small molecules, e.g., amino acids, glucose, and drugs, across membranes; symporters transport small molecules and ions in the same direction, and antiporters, in the opposite direction. Transporter superfamilies include facilitative transporters and active ATP binding cassette transporters involved in multiple-drug resistance and the targeting of antigenic peptides to MHC Class I molecules. These transporters bind to a specific ion or other molecule and undergo conformational changes in order to transfer the ion or molecule across a membrane. Transport can occur by a passive, concentration-dependent mechanism or can be linked to an energy source such as ATP hydrolysis or an ion gradient.

Ion channels, ion pumps, and transport proteins mediate the transport of molecules across cellular membranes. Symporters and antiporters regulate cytosolic pH by transporting ions and small molecules such as amino acids, glucose, and drugs. Symporters transport small molecules and ions unidirectionally, and antiporters,

binding cassette transporters which are involved in multiple-drug resistance and the targeting of antigenic peptides to MHC Class I molecules. These transporters bind to a specific ion or other molecule and undergo a conformational change in order to transfer the ion or molecule across the membrane. Transport can occur by a passive, concentration-dependent mechanism or can be linked to an energy source such as ATP hydrolysis. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland Publishing, New York, NY, pp. 523-546.)

Ion channels are formed by transmembrane proteins which create a lined passageway across the membrane through which water and ions, such as Na^+, K^+, Ca^{2+}, and Cl^-, enter and exit the cell. For example, chloride channels are involved in the regulation of the membrane electric potential as well as absorption and secretion of ions across the membrane. Chloride channels also regulate the internal pH of membrane-bound organelles.

Ion pumps are ATPases which actively maintain membrane gradients. Ion pumps are classified as P, V, or F according to their structure and function. All have one or more binding sites for ATP in their cytosolic domains. The P-class ion pumps include Ca^{2+} ATPase and Na^+/K^+ ATPase and function in transporting H^+, Na^+, K^+, and Ca^{2+} ions. P-class pumps consist of two α and two β transmembrane subunits. The V- and F-class ion pumps have similar structures and but transport only H^+. F class H^+ pumps mediate transport across the membranes of mitochondria and chloroplasts, while V-class H^+ pumps regulate acidity inside lysosomes, endosomes, and plant vacuoles.

A family of structurally related intrinsic membrane proteins known as facilitative glucose transporters catalyze the movement of glucose and other selected sugars across the plasma membrane. The proteins in this family contain a highly conserved, large transmembrane domain comprised of 12 α-helices, and several weakly conserved, cytoplasmic and exoplasmic domains (Pessin, J. E., and Bell, G.I. (1992) Annu. Rev. Physiol. 54:911-930).

Amino acid transport is mediated by Na^+ dependent amino acid transporters. These transporters are involved in gastrointestinal and renal uptake of dietary and cellular amino acids and in neuronal reuptake of neurotransmitters. Transport of cationic amino [illegible] family. Members of the CAT family share a high degree of sequence homology, and each contains 12-14 putative transmembrane domains (Ito, K. and Groudine, M. (1997) J. Biol. [illegible]

Proton-coupled, 12 membrane-spanning domain transporters such as PEPT 1 and PEPT 2 are responsible for gastrointestinal absorption and for renal reabsorbtion of peptides using an electrochemical H^+ gradient as the driving force. A heterodimeric peptide transporter, consisting of TAP 1 and TAP 2, is associated with antigen processing. Peptide antigens are transported across the membrane of the endoplasmic reticulum so

they can be presented to the major histocompatibility complex class I molecules. Each TAP protein consists of multiple hydrophobic membrane spanning segments and a highly conserved ATP-binding cassette. (Boll, M. et al. (1996) Proc. Natl. Acad. Sci. 93:284-289.)

Hormones are secreted molecules that travel through the circulation and bind to specific receptors on the surface of, or within, target cells. Although they have diverse biochemical compositions and mechanisms of action, hormones can be grouped into two categories. One category consists of small lipophilic hormones that diffuse through the plasma membrane of target cells, bind to cytosolic or nuclear receptors, and form a complex that alters gene expression. Examples of these molecules include retinoic acid, thyroxine, and the cholesterol-derived steroid hormones such as progesterone, estrogen, testosterone, cortisol, and aldosterone. The second category consists of hydrophilic hormones that function by binding to cell surface receptors that transduce signals across the plasma membrane. Examples of such hormones include amino acid derivatives such as catecholamines and peptide hormones such as glucagon, insulin, gastrin, secretin, cholecystokinin, adrenocorticotropic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, and vasopressin. (See, for example, Lodish et al. (1995) Molecular Cell Biology, Scientific American Books Inc., New York, NY, pp. 856-864.)

Neuropeptides and vasomediators (NP/VM) comprise a large family of endogenous signaling molecules. Included in this family are neuropeptides and neuropeptide hormones such as bombesin, neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, galanin, somatostatin, tachykinins, urotensin II and related

and circulatory system-borne signaling molecules such as angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides, glucagon, cholecystokinin and gastrin.

neurotransmitters and hormones, and act as catalytic enzymes in cascades. The effects of NP/VMs range from extremely brief to long-lasting. (Reviewed in Martin, C. R. et al. (1985) Endocrine Physiology, Oxford University Press, New York, NY, pp. 57-62.)

Regulatory molecules turn individual genes or groups of genes on and off in response to various inductive mechanisms of the cell or organism; act as transcription factors by determining

whether or not transcription is initiated, enhanced, or repressed; and splice transcripts as dictated in a particular cell or tissue. Although they interact with short stretches of DNA scattered throughout the entire genome, most gene expression is regulated near the site at which transcription starts or within the open reading frame of the gene being expressed. Many of the transcription factors incorporate one of a set of DNA-binding structural motifs, each of which contains either α helices or ß sheets and binds to the major groove of DNA. (Pabo, C.O. and R.T. Sauer (1992) Ann. Rev. Biochem. 61:1053-95.) Other domains of transcription factors may form crucial contacts with the DNA. In addition, accessory proteins provide important interactions which may convert a particular protein complex to an activator or a repressor or may prevent binding. (Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing Co, New York, NY pp. 401-474.)

The discovery of new human signal peptide-containing proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative disorders including cancer; inflammation; and cardiovascular, neurological, reproductive, and developmental disorders.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, proteins with signal peptides, referred to collectively as "HSPP" and individually as "HSPP-1", "HSPP-2", "HSPP-3", "HSPP-4", "HSPP-5", "HSPP-6", "HSPP-7", "HSPP-8", "HSPP-9", "HSPP-10", "HSPP-11", "HSPP-12", "HSPP-13", "HSPP-14", "HSPP-15", "HSPP-16", "HSPP-17", "HSPP-18", "HSPP-19", "HSPP-20", "HSPP-21", "HSPP-22", "HSPP-23", "HSPP-24", "HSPP-25", "HSPP-26", "HSPP-27", "HSPP-28", "HSPP-29", "HSPP-30", "HSPP-31", "HSPP-32", "HSPP-33", "HSPP-34", "HSPP-35", "HSPP-36", "HSPP-37", "HSPP-38", "HSPP-39", "HSPP-40", "HSPP-41", "HSPP-42", "HSPP-43", "HSPP-44", "HSPP-45", "HSPP-46", "HSPP-47", "HSPP-48", "HSPP-49", "HSPP-50", "HSPP-51", "HSPP-59", "HSPP-60", "HSPP-61", "HSPP-62", "HSPP-63", "HSPP-64", "HSPP-65", "HSPP-66", "HSPP-67", "HSPP-68", "HSPP-69", "HSPP-70", "HSPP-71", "HSPP-72", "HSPP-73", "HSPP-74", "HSPP-75", HSPP-76", "HSPP-77", "HSPP-78", "HSPP-79", "HSPP-87", "HSPP-88", "HSPP-89", "HSPP-90", "HSPP-91", "HSPP-92", "HSPP-93", "HSPP-94", "HSPP-95", "HSPP-96", "HSPP-97", "HSPP-98", "HSPP-99", "HSPP-100", "HSPP-

101", "HSPP-102", "HSPP-103", "HSPP-104", "HSPP-105", "HSPP-106", "HSPP-107", "HSPP-108", "HSPP-109", "HSPP-110", HSPP-111", "HSPP-112", "HSPP-113", "HSPP-114", "HSPP-115", "HSPP-116", "HSPP-117", "HSPP-118", "HSPP-119", "HSPP-120", "HSPP-121", "HSPP-122", "HSPP-123", "HSPP-124", "HSPP-125", "HSPP-126", "HSPP-127", "HSPP-128", "HSPP-129", "HSPP-130", "HSPP-131", "HSPP-132", "HSPP-133", and "HSPP-134". In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO: 28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, [illegible] ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID [illegible] NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:119, SEQ ID NO:120, SEQ ID NO:121, SEQ ID NO:122, SEQ ID NO:123, SEQ ID NO:124, SEQ ID NO:125, SEQ ID NO:126, SEQ ID NO:127, SEQ ID NO:128, SEQ ID NO:129, SEQ ID NO:130, SEQ ID

NO:131, SEQ ID NO:132, SEQ ID NO:133, SEQ ID NO:134 (SEQ ID NO:1-134), and fragments thereof.

The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-134, and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-134, and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-134, and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-134, and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-134, and fragments thereof.

The invention also provides an isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:135, SEQ ID NO:136, SEQ ID NO:137, SEQ ID NO:138, SEQ ID NO:139, SEQ ID NO:140, SEQ ID NO:141, SEQ ID NO:142, SEQ ID NO:143, SEQ ID NO:144, SEQ ID NO:145, SEQ ID NO:146, SEQ ID NO:147, SEQ ID NO:148, SEQ ID NO:149, SEQ ID NO:150, SEQ ID NO:151, SEQ ID NO:152, SEQ ID NO:153, SEQ ID NO:154, SEQ ID NO:155, SEQ ID NO:156, SEQ ID NO:157, SEQ ID NO:158, SEQ ID NO:159, SEQ ID NO:160, SEQ ID NO:161, SEQ ID NO:162, SEQ ID NO:163, SEQ ID NO:164, SEQ ID NO:165, SEQ ID NO:166, SEQ ID NO:167, SEQ ID NO:168, SEQ ID NO:169, SEQ ID NO:170, SEQ ID NO:171, SEQ ID NO:172, SEQ ID NO:173, SEQ ID NO:174, SEQ ID NO:175, SEQ ID NO:176, SEQ ID NO:177, SEQ ID NO:178, SEQ ID NO:179, SEQ ID NO:180, SEQ ID NO:181, SEQ ID NO:182, SEQ ID NO:183, SEQ ID NO:184, SEQ ID NO:185, SEQ ID NO:186, SEQ ID NO:187, SEQ ID NO:188, SEQ ID NO:189, SEQ ID NO:190, SEQ ID NO:191, SEQ ID NO:192, SEQ ID NO:193, SEQ ID NO:194, SEQ ID NO:195, SEQ ID

NO:196, SEQ ID NO:197, SEQ ID NO:198, SEQ ID NO:199, SEQ ID NO:200, SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203, SEQ ID NO:204, SEQ ID NO:205, SEQ ID NO:206, SEQ ID NO:207, SEQ ID NO:208, SEQ ID NO:209, SEQ ID NO:210, SEQ ID NO:211, SEQ ID NO:212, SEQ ID NO:213, SEQ ID NO:214, SEQ ID NO:215, SEQ ID NO:216, SEQ ID NO:217, SEQ ID NO:218, SEQ ID NO:219, SEQ ID NO:220, SEQ ID NO:221, SEQ ID NO:222, SEQ ID NO:223, SEQ ID NO:224, SEQ ID NO:225, SEQ ID NO:226, SEQ ID NO:227, SEQ ID NO:228, SEQ ID NO:229, SEQ ID NO:230, SEQ ID NO:231, SEQ ID NO:232, SEQ ID NO:233, SEQ ID NO:234, SEQ ID NO:235, SEQ ID NO:236, SEQ ID NO:237, SEQ ID NO:238, SEQ ID NO:239, SEQ ID NO:240, SEQ ID NO:241, SEQ ID NO:242, SEQ ID NO:243, SEQ ID NO:244, SEQ ID NO:245, SEQ ID NO:246, SEQ ID NO:247, SEQ ID NO:248, SEQ ID NO:249, SEQ ID NO:250, SEQ ID NO:251, SEQ ID NO:252, SEQ ID NO:253, SEQ ID NO:254, SEQ ID NO:255, SEQ ID NO:256, SEQ ID NO:257, SEQ ID NO:258, SEQ ID NO:259, SEQ ID NO:260, SEQ ID NO:261, SEQ ID NO:262, SEQ ID NO:263, SEQ ID NO:264, SEQ ID NO:265, SEQ ID NO:266, SEQ ID NO:267, SEQ ID NO:268 (SEQ ID NO:135-268), and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:135-268, and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:135-268, and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of (a) hybridizing the

sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence

the polynucleotide prior to hybridization.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-134, and fragments thereof. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-134, and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-134, and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of HSPP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-134, and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of HSPP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-134, and fragments thereof.

BRIEF DESCRIPTION OF THE TABLE

Table 1 shows nucleotide and polypeptide sequence identification numbers (SEQ ID NO), clone identification numbers (clone ID), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding HSPP.

Table 2 shows features of each polypeptide sequence, including predicted signal peptide sequences, and methods and algorithms used for identification of HSPP.

Table 3 shows the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis, diseases, disorders, or conditions associated with these tissues, and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which Incyte cDNA clones encoding HSPP were isolated.

Table 5 shows the programs, their descriptions, references, and threshold parameters used to analyze HSPP.

Table 6 shows the regions of the full-length nucleotide sequences of HSPP to which cDNA fragments of Table 1 correspond.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"HSPP" refers to the amino acid sequences of substantially purified HSPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which, when bound to HSPP, increases or prolongs the duration of the effect of HSPP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of HSPP.

An "allelic variant" is an alternative form of the gene encoding HSPP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HSPP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same as HSPP or a polypeptide with at least one functional characteristic of HSPP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HSPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HSPP. The encoded

amino acid residues which produce a silent change and result in a functionally equivalent HSPP. Deliberate amino acid substitutions may be made on the basis of similarity in

of the residues, as long as the biological or immunological activity of HSPP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine,

isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments," "immunogenic fragments," or "antigenic fragments" refer to fragments of HSPP which are preferably at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity or immunological activity of HSPP. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which, when bound to HSPP, decreases the amount or the duration of the effect of the biological or immunological activity of HSPP. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of HSPP.

The term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind HSPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete

with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

The term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HSPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding HSPP or fragments of HSPP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence"refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW Fragment Assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding HSPP, by northern analysis is indicative of the presence of nucleic acids encoding HSPP in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding HSPP.

A "deletion"refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word

identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern [illegible] blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require

that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Madison WI) which creates alignments between two or more sequences according to methods selected by the user, e.g., the clustal method. (See, e.g., Higgins, D.G. and P.M. Sharp (1988) Gene 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) Methods Enzymol. 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs) are linear microchromosomes which [illegible] elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions [illegible] altered so [illegible] more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary

bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" or "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of HSPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of HSPP.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic

to those nucleic acid sequences which, comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:135-268, for example, as distinct from any other sequence in the same genome. For example, a fragment of SEQ ID NO:135-268 is useful in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:135-268 from related polynucleotide sequences. A fragment of SEQ ID NO:135-268 is at least about 15-20 nucleotides in length. The precise length of the fragment of SEQ ID NO:135-268 and the region of SEQ ID NO:135-268 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based

on the intended purpose for the fragment. In some cases, a fragment, when translated, would produce polypeptides retaining some functional characteristic, e.g., antigenicity, or structural domain characteristic, e.g., ATP-binding site, of the full-length polypeptide.

The terms "operably associated" or "operably linked" refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the translation of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the sequence encoding the polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. "Oligonucleotide" is substantially equivalent to the terms "amplimer," "primer," "oligomer," and "probe," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

The term "sample" is used in its broadest sense. A sample suspected of containing

fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the

presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "stringent conditions" refers to conditions which permit hybridization between polynucleotides and the claimed polynucleotides. Stringent conditions can be defined by salt concentration, the concentration of organic solvent, e.g., formamide, temperature, and other conditions well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host

transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of HSPP polypeptides refers to an amino acid sequence that is altered by one or more amino acid residues. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g.,

replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to HSPP. This definition may also include, for example, "allelic" (as defined above), "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

THE INVENTION

The invention is based on the discovery of new human signal peptide-containing [illegible] for the diagnosis, treatment, or prevention of cell proliferative disorders including cancer; inflammation; and cardiovascular, neurological, reproductive, and developmental disorders.

Table 1 lists t[illegible] Clones used [illegible] derive full length nucleotide sequences encoding HSPP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NO) of the amino acid and nucleic acid sequences, respectively. Column 3 shows the Clone ID of the Incyte Clone in which nucleic acids encoding each HSPP were identified, and column 4, the cDNA libraries from which these clones were isolated. Column 5

shows Incyte clones, their corresponding cDNA libraries, and shotgun sequences. The clones and shotgun sequences are part of the consensus nucleotide sequence of each HSPP and are useful as fragments in hybridization technologies.

Table 6 shows the regions of the full-length nucleotide sequences of HSPP to which cDNA fragments of Table 1 correspond. Column 1 lists nucleotide sequence identifiers and column 2 shows the clone ID of the Incyte clone in which nucleic acids encoding each HSPP were identified. Column 3 shows Incyte clones and shotgun sequences which are part of the consensus nucleotide sequence of each HSPP and are useful as fragments in hybridization technologies. Column 4 lists the starting nucleotide position and column 5 the ending nucleotide position of the region of the full-length HSPP to which the cDNA fragment corresponds.

The columns of Table 2 show various properties of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3, potential phosphorylation sites; column 4, potential glycosylation sites; column 5, the amino acid residues comprising signature sequences and motifs; column 6, the identity of each protein; and column 7, analytical methods used to identify each HSPP as a signal peptide-containing protein. Note that in column 5, the first line of each cell lists the amino acid residues comprising predicted signal peptide sequences. Additional identifying motifs or signatures are also listed in column 5. Of particular note is the presence of a glycosyl hydrolase family 9 active site signature in SEQ ID NO:126, a ribosomal protein S18 signature in SEQ ID NO:127, an adrenodoxin family iron-sulfur binding region signature and a cytochrome c family heme-binding site signature in SEQ ID NO:132, and a urotensin II signature sequence in SEQ ID NO:

Using BLAST, SEQ ID NO:68 (HSPP-68) has been identified as a TWIK-related acid-sensitive K^+ channel and SEQ ID NO:92 (HSPP-92) has been identified as a tyrosine-specific protein phosphatase. The tyrosine-specific protein phosphatases signature in SEQ ID NO:92 (HSPP-92) from about V328 through about F340 (including the putative active site cysteine residue at C330) was identified using BLOCKS and PRINTS. Also of note is the identification of SEQ ID NO:66 (HSPP-66) as a steroid binding protein using BLAST.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding HSPP. The first column of Table 3 lists the nucleotide sequence identifiers. The second column lists tissue categories which express HSPP as a fraction of total tissue categories expressing HSPP. The third column lists the diseases, disorders, or conditions associated with those tissues expressing HSPP. The fourth column lists the vectors used to subclone the cDNA library. Of particular note is the expression of SEQ ID NO:200, SEQ ID NO:203, and SEQ ID NO:225 in lung tissues; the expression of SEQ ID NO:212, SEQ ID NO:216, and SEQ ID NO:220 in reproductive tissues; the expression of SEQ ID NO:223 in cancerous tissues; the expression of SEQ ID NO:232 in gastrointestinal tissue, specifically the small intestine or colon (fifteen out of sixteen (93.8%) cDNA libraries); and the expression of SEQ ID NO:224 in cancerous and proliferating tissues. Also of particular interest is the tissue-specific expression of SEQ ID NO:252 and SEQ ID NO:257. SEQ ID NO:252 is derived from OVARTUT01, an ovarian tumor cDNA library and is exclusively expressed in reproductive tumor tissue. SEQ ID NO:257 is derived from THP1AZT01, a 5-aza-2'-deoxycytidine treated human promonocyte cDNA library and is exclusively expressed in hematopoietic tissue.

The following fragments of the nucleotide sequences encoding HSPP are useful in hybridization or amplification technologies to identify SEQ ID NO:135-268 and to distinguish between SEQ ID NO:135-268 and related polynucleotide sequences. The useful fragments are the fragment of SEQ ID NO:230 from about nucleotide 75 to about nucleotide 104; the fragment of SEQ ID NO:231 from about nucleotide 210 to about nucleotide 239; the fragment of SEQ ID NO:232 from about nucleotide 157 to about [illegible] nucleotide 297; the fragment of SEQ ID NO:234 from about nucleotide 160 to about nucleotide 186; the fragment of SEQ ID NO:235 from about nucleotide 201 to about nucleotide 230, the fragment of SEQ ID NO:236 from about nucleotide [illegible] to about nucleotide 194; the fragment of SEQ ID NO:237 from about nucleotide 366 to about nucleotide 395; the fragment of SEQ ID NO:238 from about nucleotide 714 to about nucleotide 743; the fragment of SEQ ID NO:239 from about nucleotide 1731 to about nucleotide 1760; the fragment of SEQ ID NO:240 from about nucleotide 419 to about nucleotide 448; the fragment of SEQ ID NO:241 from about nucleotide 494 to about

nucleotide 523; the fragment of SEQ ID NO:242 from about nucleotide 100 to about nucleotide 129; the fragment of SEQ ID NO:243 from about nucleotide 104 to about nucleotide 133; the fragment of SEQ ID NO:244 from about nucleotide 136 to about nucleotide 165; the fragment of SEQ ID NO:245 from about nucleotide 140 to about nucleotide 169; the fragment of SEQ ID NO:246 from about nucleotide 125 to about nucleotide 154; the fragment of SEQ ID NO:247 from about nucleotide 687 to about nucleotide 758; the fragment of SEQ ID NO:248 from about nucleotide 327 to about nucleotide 398; the fragment of SEQ ID NO:249 from about nucleotide 741 to about nucleotide 785; the fragment of SEQ ID NO:250 from about nucleotide 184 to about nucleotide 255; the fragment of SEQ ID NO:251 from about nucleotide 165 to about nucleotide 242; the fragment of SEQ ID NO:252 from about nucleotide 271 to about nucleotide 342; the fragment of SEQ ID NO:253 from about nucleotide 1081 to about nucleotide 1152; the fragment of SEQ ID NO:254 from about nucleotide 781 to about nucleotide 852; the fragment of SEQ ID NO:255 from about nucleotide 620 to about nucleotide 691; the fragment of SEQ ID NO:256 from about nucleotide 872 to about nucleotide 916; the fragment of SEQ ID NO:257 from about nucleotide 242 to about nucleotide 313; the fragment of SEQ ID NO:258 from about nucleotide 595 to about nucleotide 648; the fragment of SEQ ID NO:259 from about nucleotide 163 to about nucleotide 216; the fragment of SEQ ID NO:260 from about nucleotide 244 to about nucleotide 315; the fragment of SEQ ID NO:261 from about nucleotide 75 to about nucleotide 128; the fragment of SEQ ID NO:262 from about nucleotide 650 to about nucleotide 703; the fragment of SEQ ID NO:263 from about nucleotide 143 to about nucleotide 214; the fragment of SEQ ID NO:264 from about nucleotide 434 to about [illegible] nucleotide 271; the fragment of SEQ ID NO:266 from about nucleotide 89 to about nucleotide 145; the fragment of SEQ ID NO:267 from about nucleotide 198 to about nucleotide 254; and the fragment of SEQ ID NO:268 from about nucleotide 10 to about nucleotide 54.

The invention also encompasses HSPP variants. A preferred HSPP variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the HSPP amino acid sequence, and which contains at least one functional or structural characteristic of HSPP.

The invention also encompasses polynucleotides which encode HSPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:135-268, which encodes HSPP.

The invention also encompasses a variant of a polynucleotide sequence encoding HSPP. In particular, such a variant polynucleotide sequence will have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding HSPP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:135-268 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:135-268. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HSPP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding HSPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring HSPP, and all such variations are to be considered as being specifically disclosed.

Alt[illegible] capable of hybridizing to the nucleotide sequence of the naturally occurring HSPP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HSPP or its derivatives possessing a substantially different [illegible], inclu[illegible]. Codons ma[illegible] elected t increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HSPP and its derivatives without altering the encoded amino acid sequences include the

production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode HSPP and HSPP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding HSPP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:135-268 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art.

needed. In a preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and [illegible] DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42°C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50 % formamide, and 200 μg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps which follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25°C, more preferably of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the Hamilton MICROLAB 2200 (Hamilton, Reno NV), Peltier Thermal Cycler 200 (PTC200; MJ Research, Watertown MA) and the ABI CATALYST 800 [illegible] sequencing systems (Perkin-Elmer) or the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA). The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short [illegible], John [illegible]ley & [illegible] New York NY, [illegible] Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding HSPP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect

upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-306). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which [illegible] which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output light intensity may be converted to electrical signal using appropriate software (e.g.,

GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode HSPP may be cloned in recombinant DNA molecules that direct expression of HSPP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express HSPP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HSPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

In another embodiment, sequences encoding HSPP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, and Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232.) Alternatively, HSPP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various

Automated synthesis may be achieved using the ABI 431A Peptide Synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of HSPP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g, Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid

analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active HSPP, the nucleotide sequences encoding HSPP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HSPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HSPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HSPP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding HSPP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA

J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

xpressi /host utilized t

sequences encoding HSPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral

expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding HSPP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding HSPP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or pSPORT1 plasmid (Life Technologies). Ligation of sequences encoding HSPP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of HSPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of HSPP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of HSPP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Grant et al. (1987) Methods Enzymol.

Plant systems may also be used for expression of HSPP. Transcription of sequences encoding HSPP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated

transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding HSPP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses HSPP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of HSPP in cell lines is preferred. For example, sequences encoding HSPP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, [illegible] cell introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. [illegible] to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ or *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to

the aminoglycosides, neomycin and G-418; and *als* or *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HSPP is inserted within a marker gene sequence, transformed cells containing sequences encoding HSPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HSPP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding HSPP and that express HSPP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques [illegible] quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of HSPP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techni[illegible]d immun[illegible]t assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HSPP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al.

(1990) Serological Methods, a Laboratory Manual, APS Press, St Paul MN, Sect. IV; Coligan, J. E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding HSPP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HSPP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding HSPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode HSPP may be [illegible] or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such [illegible] glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK,

HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Manassas, VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding HSPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric HSPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HSPP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HSPP encoding sequence and the heterologous protein sequence, so that HSPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

d HSPP

achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract systems (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of labeled amino acid precursor, preferably ^{35}S-methionine.

Fragments of HSPP may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A Peptide

Synthesizer (Perkin-Elmer). Various fragments of HSPP may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of HSPP and signal peptide sequences. In addition, chemical and structural similarity, in the context of sequences and motifs, exists between HSPP-66 and prostatic steriod-binding C3 precursor from rat (GI 206453); between HSPP-68 and TWIK-related acid-sensitive K+channel from human (GI 2465542); and between HSPP-92 and tyrosine specific protein phosphatases (PROSITE PDOC00323). In addition, the expression of HSPP is closely associated with proliferative, cancerous, inflamed, cardiovascular, nervous, reproductive, hematopoietic/immune, and developmental tissue. Therefore, HSPP appears to play a role in cell proliferative disorders including cancer; inflammation; and cardiovascular, neurological, reproductive, and developmental disorders. In the treatment of cell proliferative disorders including cancer; inflammation; and cardiovascular, neurological, reproductive, and developmental disorders associated with increased HSPP expression or activity, it is desirable to decrease the expression or activity of HSPP. In the treatment of the above conditions associated with decreased HSPP expression or activity, it is desirable to increase the expression or activity of HSPP.

Therefore, in one embodiment, HSPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HSPP. Examples of such disorders include, but are not limited to, [illegible]sclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, [illegible]ratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; inflammatory disorders, such as acquired immunodeficiency syndrome (AIDS), Addison's

disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; cardiovascular disorders including disorders of the blood vessels such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, and vascular tumors; disorders of the heart such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, and congenital

lungs tele

pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease,

hyser bronchial asthma, bro isis, bacte

viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary

hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, and pleural tumors; neurological disorders such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; reproductive disorders such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous

n

hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physi ,

nce f th p

Peyronie's disease, carcinoma of the male breast, and gynecomastia; and developmental disorders, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental

retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss.

In another embodiment, a vector capable of expressing HSPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HSPP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified HSPP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HSPP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of HSPP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of HSPP including, but not limited to, those listed above.

In a further embodiment, an antagonist of HSPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HSPP. Examples of such disorders include, but are not limited to, those described above. In one aspect, an antibody which specifically binds HSPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express HSPP.

'he ic

polynucleotide encoding HSPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HSPP including, but not limited to, those described above.

the proteins, a nists, antibodies, agoni

complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act

synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of HSPP may be produced using methods which are generally known in the art. In particular, purified HSPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HSPP. Antibodies to HSPP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HSPP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HSPP have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid [illegible] naturally occurring molecule. Short stretches of HSPP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

[illegible] to HSPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42;

Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HSPP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for HSPP may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

anti vin desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex [illegible] HSPP specific a[illegible], monoclon[illegible] immunoassay utilizing monoclonal antibodies reactive to two non-interfering HSPP epitopes is preferred, but a competitive binding assay may also be employed (Pound supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for HSPP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of HSPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple HSPP epitopes, represents the average affinity, or avidity, of the antibodies for HSPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular HSPP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the HSPP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of HSPP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons. New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is preferred for use in procedures requiring precipitation of HSPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and

nd Co[illegible]n et al. supra.)

In another embodiment of the invention, the polynucleotides encoding HSPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, [illegible] of the polynucleotide encoding HSPP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding HSPP. Thus, complementary molecules or fragments may be used to modulate HSPP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and

sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding HSPP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding HSPP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding HSPP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding HSPP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding HSPP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases,

DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco , pp. 163-177.) A complementary sequence or antisense molecule may also be esigne prevent from ding ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by

endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding HSPP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding HSPP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather

inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine,

[illegible] by endogeno[illegible]

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or

by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of HSPP, antibodies to HSPP, and mimetics, agonists, antagonists, or inhibitors of HSPP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as t[illegible]s, liqui[illegible]gels, sy[illegible]p[illegible]s, a[illegible]f[illegible] ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient a[illegible] processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be

added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection

[illegible] inc[illegible]

as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid [illegible]les, o[illegible]es. No[illegible] amin[illegible] polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of HSPP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially

rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example HSPP or fragments thereof, antibodies of HSPP, and agonists, antagonists or inhibitors of HSPP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically

effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind HSPP may be used for the diagnosis of disorders characterized by expression of HSPP, or in assays to monitor patients being treated with HSPP or agonists, antagonists, or inhibitors of HSPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HSPP include methods which utilize the antibody and a label to detect HSPP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled

by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring HSPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of HSPP expression. Normal or standard values for HSPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HSPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of HSPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding HSPP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of HSPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of HSPP, and to monitor regulation of HSPP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HSPP or closely related molecules may be used to identify nucleic acid sequences which encode HSPP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the

ate, e.

determine whether the probe identifies only naturally occurring sequences encoding HSPP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and should

y to a enco

The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:135-268 or from genomic sequences including promoters, enhancers, and introns of the HSPP gene.

Means for producing specific hybridization probes for DNAs encoding HSPP include the cloning of polynucleotide sequences encoding HSPP or HSPP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding HSPP may be used for the diagnosis of disorders associated with expression of HSPP. Examples of such disorders include, but are not limited to, cell proliferative disorders such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; inflammatory disorders, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, [illegible]ema lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia [illegible]is, myoc[illegible]ma[illegible]n, oste[illegible]is, osteoporosis, p[illegible]eatitis polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic,

protozoal, and helminthic infections, and trauma; cardiovascular disorders including disorders of the blood vessels such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, and vascular tumors; disorders of the heart such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, and congenital heart disease; and disorders of the lungs such as congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, and pleural tumors; neurological disorders such as epilepsy, ischemic cerebrovascular disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal

hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; reproductive disorders such as disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, carcinoma of the male breast, and gynecomastia; and developmental disorders, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, [illegible] is, [illegible] and [illegible]sc loss. The polynucleotide sequences encoding HSPP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered HSPP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HSPP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HSPP may be labeled by standard methods and added

to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HSPP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of HSPP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding HSPP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

[illegible] ipt (c

under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding HSPP may involve the use of PCR. These oligomers may be chemically

synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding HSPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding HSPP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of HSPP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application [illegible], R.A. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding HSPP [illegible] used to [illegible]ate hybridiza[illegible]es use[illegible] in mappi[illegible] naturally oc[illegible] genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries.

(See, e.g., Harrington, J.J. et al. (1997) Nat Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding HSPP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, normal

In another embodiment of the invention, HSPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between HSPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen,

et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with HSPP, or fragments thereof, and washed. Bound HSPP is then detected by methods well known in the art. Purified HSPP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding HSPP specifically compete with a test compound for binding HSPP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HSPP.

In additional embodiments, the nucleotide sequences which encode HSPP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all applications, patents, and publications, mentioned above and below, in particular US Ser. No. 60/090,762, US Ser. No. 60/094,983, US Ser. No. 60/102,686, and US Ser. No. 60/112,129, are hereby expressly incorporated by reference.

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homog ized an ysed in g nidinium isothiocyanate, while others were h mo mi suitab re o TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Valencia CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), pSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

Plasmids were recovered from host cells by in vivo excision, using the UNIZAP vector system (Stratagene) or cell lysis. Plasmids were purified using at least one of the following: a MAGIC or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the REAL Prep 96 plasmid kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4° C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a Fluoroskan II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

The cDNAs were prepared for sequencing using the ABI CATALYST 800 (Perkin-Elmer) or the HYDRA microdispenser (Robbins Scientific) or MICROLAB 2200 (Hamilton) systems in combination with the PTC-200 thermal cyclers (MJ Research). The cDNAs were sequenced using the ABI PRISM 373 or 377 sequencing systems (Perkin-Elmer) and standard ABI protocols, base calling software, and kits. In one alternative, cDNAs were sequenced using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics). In another alternative, the cDNAs were amplified and sequenced using the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). In yet another alternative, cDNAs were sequenced using solutions and dyes from Amersham Pharmacia Biotech. Reading frames for the ESTs were determined using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA, extension, and shotgun sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the software programs, descriptions, references, and threshold parameters used. The first column of [illegible] provides a brief description thereof, the third column presents the references which are incorporated by reference herein, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between t[illegible]equer[illegible] the probability t[illegible] the hom[illegible] analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR).

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based

on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS to acquire annotation, using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probalistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Cur. Opin. Str. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:135-268. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

or related molecules in nucleotide databases such as GenBank or LIFESEQ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybr ...ions. In ac ...n, the sensitivity of the computer search can be modified to ... ...ine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of Northern analyses are reported as a percentage distribution of libraries in which the transcript encoding HSPP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation/trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Extension of HSPP Encoding Polynucleotides

Full length nucleic acid sequences of SEQ ID NOs:135-229 were produced by extension of the component fragments described in Table 1, column 5, using oligonucleotide primers based on these fragments. For each nucleic acid sequence, one primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were

another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68 °C to about 72 °C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR™ kit (The Perkin-Elmer Corp., Norwalk, CT) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the PTC-200 thermal cycler (MJ Research, Inc., Watertown, MA), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat steps 4 through 6 for an additional 15 cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat steps 8 through 10 for an additional 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK™ (QIAGEN Inc.), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent E. coli cells (in appropriate media) were transformed with 3 µl of ligation mixture and cultured in 80 µl of SOC medium. (See, e.g., Sambrook, Supra, Appendix A, p. 2.) After incubation for one hour at 37°C, the E. coli mixture was plated on Bc[illegible] carbenicillin (2x carb). The following day, several colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/2x carb medium placed in an individual well of an appropriate commercially-available sterile 96-well microtiter plate. The following day, 5 µl of [illegible] strat[illegible] plate and, after dilution 1:10 with water, 5 μl from each sample was transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2 through 4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

The full length nucleic acid sequences of SEQ ID NO:230-268 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.) with a 96-well DNA Engine. The reaction mix contained 2 μl of template, 2.5 pmol of each primer, 2.5 μl of 10x reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as

follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified under the same conditions as described above. Samples were diluted with 20% dimethysulphoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:135-268 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:135-268 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba1, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT-AR film (Eastman Kodak, Rochester NY) is exposed to the blots to film for several hours, hybridization

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of

complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the HSPP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring HSPP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of HSPP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the HSPP-encoding transcript.

IX. Expression of HSPP

Expression and purification of HSPP is achieved using bacterial or virus-based expression systems. For expression of HSPP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria

express HSPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of HSPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding HSPP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E. K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, HSPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from HSPP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch 10 and 16). Purified HSPP obtained by these

V. Demonstration of HSPP Activity

HSPP-68

HSPP-68 activity is measured by determining the potassium current using voltage clamp analysis on single Xenopus laevis oocytes injected with HSPP-68 cRNA. HSPP-68 cRNA is synthesized in vitro from linearized HSPP-68 encoding plasmids using the T7

RNA polymerase and injected into oocytes.. Injected oocytes are used two to four days after injection. In a 0.3 ml perfusion chamber, a single oocyte is impaled with two standard microelectrodes (1-2.5 MΩ) filled with 3 M KCl. The oocyte is maintained under voltage clamp by using a Dagan TEV 200 amplifier, in buffer containing 96 mM NaCl, 2 mM KCl, 1.8 mM $CaCl_2$, 2 mM $MgCl_2$, 5 mM HEPES, pH 7.4 with NaOH. Stimulation of the preparation, data acquisition, and analysis is performed using a computer. All experiments are performed at room temperature (21-22 °C). Following a depolarizing pulse, the characteristics of the resulting potassium current are measured via the recording electrode. The amount of potassium current that flows in response to a unit depolarization is proportional to the activity of HSPP-68 in the cell. (Duprat, F. et al. (1997) EMBO J. 16:5464-5471.)

HSPP-92

HSPP-92 protein phosphatase activity is measured by the hydrolysis of P-nitrophenyl phosphate (PNPP). HSPP-92 is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% b-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH and the increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of PP in the assay. (Diamond R.H. et al (1994) Mol Cell Biol 14:3752-62.)

Alternatively, HSPP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HSPP, washed, and any wells with labeled HSPP complex are assayed. Data

number, affinity, and association of HSPP with the candidate molecules.

Alternatively, an assay for HSPP activity measures the expression of HSPP on the cell surface. cDNA encoding HSPP is subcloned into an appropriate mammalian expression vector suitable for high levels of DNA expression. The resulting construct is transfected into a nonhuman cell line such as NIH3T3. Cell surface proteins are labeled with biotin using methods known in the art. Immunoprecipitations are performed using HSPP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to

unlabeled immunoprecipitant is proportional to the amount of HSPP expressed on the cell surface.

Alternatively, an assay for HSPP activity measures the amount of HSPP in secretory, membrane-bound organelles. Transfected cells as described above are harvested and lysed. The lysate is fractionated using methods known to those of skill in the art, for example, sucrose gradient ultracentrifugation. Such methods allow the isolation of subcellular components such as the Golgi apparatus, ER, small membrane-bound vesicles, and other secretory organelles. Immunoprecipitations from fractionated and total cell lysates are performed using HSPP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The concentration of HSPP in secretory organelles relative to HSPP in total cell lysate is proportional to the amount of HSPP in transit through the secretory pathway.

XI. Functional Assays

HSPP function is assessed by expressing the sequences encoding HSPP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression [illegible]. Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP, and to evaluate [illegible] perties for example, their apoptotic state. FCM detects and quantifies [illegible] fluorescent molecules that diagnose events [illegible] or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in

expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of HSPP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding HSPP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding HSPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of HSPP Specific Antibodies

HSPP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the HSPP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the

1995, supra, ch. 11.)

Typically, oligopeptides 15 residues in length are synthesized using an ABI 431A Peptide Synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-[illegible] MO) using N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic,

blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring HSPP Using Specific Antibodies

Naturally occurring or recombinant HSPP is substantially purified by immunoaffinity chromatography using antibodies specific for HSPP. An immunoaffinity column is constructed by covalently coupling anti-HSPP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HSPP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HSPP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HSPP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HSPP is collected.

XIV. Identification of Molecules Which Interact with HSPP

HSPP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HSPP, washed, and any wells with labeled HSPP complex are assayed. Data obtained using different concentrations of HSPP are used to calculate values for the number, affinity, and association of HSPP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and

specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

TABLE 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone	Library	Fragments
1		44353	MPHGNOT03	443531H1 (MPHGNOT03), 1406807F6 (LATRTUT02), 443531T6 (MPHGNOT03), SBBA00451F1, SBBA00676F1
2		63286	NEUTGMT01	632860H1 (NEUTGMT01), 784715R3 (PROSNOT05), 509590H1 (MPHGNOT03)
3		67001	CRBLNOT01	670010H1 (CRBLNOT01), 669971R1 (CRBLNOT01), 1553045F1 (BLADTUT04)
4		7264	SYNOOAT01	726498H1 (SYNOOAT01), 726498R6 (SYNOOAT01), 866599R3 (BRAITUT03)
5		79506	OVARNOT03	795064H1 (OVARNOT03), 4339458H1 (BRAUNOT02), 937605R3 (CERVNOT01), 2381151F6 (ISLTNOT01), 1466346F6 (PANCTUT02)
6		92492	BRAINOT04	924925H1 (BRAINOT04), 3268330H1 (BRAINOT20), 759120R3 (BRAITUT02)
7		96239	BRSTTUT03	962390H1 (BRSTTUT03), 1907958F6 (CONNTUT01), 023569F1 (ADENINB01), 167282F1 (LIVRNOT01), 1309211F1 (COLNFET02), SAUA00696F1, SAUA02860F1
8		12594	MENITUT03	1259405H1 (MENITUT03), 2472425H1 (THP1NOT03), 774303R1 (COLNNOT05), 1520779F1 (BLADTUT04), 1693833F6 (COLNNOT23), 1831858T6.comp (THP1AZT01), 1527737T6.comp (UCMCL5T01)
9	143	12973	BRSTNOT07	1297384H1 (BRSTNOT07), 1269310F6 (BRAINOT09), 1457367F1 (COLNFET02), 415587R1 (BRSTNOT01), SANA02967F1
10		12996	BRSTNOT07	1299627H1 (BRSTNOT07), 1359140F6 (LUNGNOT09), 1349224F1 (LATRTUT02), SBAA01431F1, SBAA02909F1, SBAA01156F1
11		13060	PLACNOT02	1306026H1 (PLACNOT02), 1464088R6 (PANCNOT04), SBAA02496F1, SBAA04305F1
12		13162	BLADTUT02	1316219H1 (BLADTUT02), 2458603F6 (ENDANOT01), 2504756T6 (CONUTUT01)
13		13290	PANCNOT07	1329031H1 (PANCNOT07), 1329031T6 (PANCNOT07), 1329031F6 (PANCNOT07)

TABLE 1 (cont.)

Protein SEQ ID NO:	Nucl SEQ	e O:	Clone	Library	Fragments
14			14830	CORPNOT02	1483050H1 (CORPNOT02), 855049H1 (NGANNOT01), 077017F1 (SYNORAB01), 1483050F6 (CORPNOT02), 1480024T6 (CORPNOT02), 1483050T6 (CORPNOT02), 759486R1 (BRAITUT02)
15			15141	PANCTUT01	1514160H1 (PANCTUT01), 1866765T7 (SKINBIT01), 782676R1 (MYOMNOT01), 008055X4 (HMC1NOT01), 008055X5 (HMC1NOT01), 1866765F6 (SKINBIT01), SAOA03127F1
16			16034	LUNGNOT15	1603403H1 (LUNGNOT15), 372910F1 (LUNGNOT02), 733299R7 (LUNGNOT03)
17	151		16523	PROSTUT08	1652303H1 (PROSTUT08), 1671806H1 (BLADNOT05), 1341743T1 (COLNTUT03), 3803812H1 (BLADTUT03), 1878546F6 (LEUKNOT03), 1428640F1 (SINTBST01), 2058609R6 (OVARNOT03), 1331621F1 (PANCNOT07), 1306331T1 (PLACNOT02)
18			16933	COLNNOT23	1693358H1 (COLNNOT23), 2498265H1 (ADRETUT05), 1867125F6 (SKINBIT01), 1693358T6 (COLNNOT23), 2245848R6 (HIPONON02)
19			17077	DUODNOT02	1707711H1 (DUODNOT02), 1484609T1 (CORPNOT02), 1707711F6 (DUODNOT02), 1267959F1 (BRAINOT09), 1484609F1 (CORPNOT02), SAJA00930F1, SAJA01300R1, SAJA00999R1
20			173873	COLNNOT22	1738735H1 (COLNNOT22), SAJA00944R1, SAJA00137F1, SAJA03629F1
21			17491	STOMTUT02	1749147H1 (STOMTUT02), 1749147F6 (STOMTUT02), 1749147T6 (STOMTUT02)
22	156		18177	PROSNOT20	1817722H1 (PROSNOT20), 2011085H1 (TESTNOT03)
23			18312	THP1AZT01	1831290H1 (THP1AZT01), 3473958H1 (LUNGNOT27), 1972268F6 (UCMCL5T01), 1301277F1 (BRSTNOT07), 1521574F1 (BLADTUT04), 1561690T6 (SPLNNOT04), 891461R1 (STOMTUT01)

TABLE 1 (cont.)

Protein SEQ ID NO:	Nuc SEQ de NO:	Clone	Library	Fragments
24	15	18314	THP1AZT01	1831477H1 (THP1AZT01), 1582867H1 (DUODNOT01), 1336769T1 (COLNNOT13), 1933092H1 (COLNNOT16), 1519909F1 (BLADTUT04), 1220946H1 (NEUTGMT01), 809556T1 (LUNGNOT04), 1217559T1 (NEUTGMT01), 1309225F1 (COLNFET02)
25		18416	COLNNOT07	1841607H1 (COLNNOT07), SBHA03588F1
26		18523	LUNGFET03	1852391H1 (LUNGFET03), 734140H1 (TONSNOT01), 1852391F6 (LUNGFET03)
27		18545	HNT3AZT01	1854555H1 (HNT3AZT01), 2511711H1 (CONUTUT01), 782453R1 (MYOMNOT01), 1854555F6 (HNT3AZT01), 1840675T6 (COLNNOT07), 2109736H1 (BRAITUT03)
28		18557	PROSNOT18	1855755H1 (PROSNOT18), 3040236H1 (BRSTNOT16), 1283207F1 (COLNNOT16), 833763T1 (PROSNOT07), 1920926R6 (BRSTTUT01)
29	16	18614	PROSNOT19	1861434H1 (PROSNOT19), 980291R1 (TONGTUT01), 1861434T6 (PROSNOT19), SARA01525F1, SARA02548F1
30		18723	LEUKNOT02	1872334H1 (LEUKNOT02), 1872334F6 (LEUKNOT02), SBGA03684F1
31		18772	LEUKNOT03	1877230H1 (LEUKNOT03), 2519841H1 (BRAITUT21), 1877230T6 (LEUKNOT03), 1254693F1 (LUNGFET03), 077020R1 (SYNORAB01), 1232336F1 (LUNGFET03), 1004952R6 (BRSTNOT03), SARA01879F1, SARA02654F1
32		18778	LEUKNOT03	1877885H1 (LEUKNOT03), 508020F1 (TMLR3DT01), 2751126R6 (THP1AZS08), SARA02571F1
33	67	18892	BLADTUT07	1889269H1 (BLADTUT07), 1915551H1 (PROSTUT04), 629493X12 (KIDNNOT05), 1441289F1 (THYRNOT03), 1215274X34F1 (BRSTTUT01), 1818447F6 (PROSNOT20), 1208463R1 (BRSTNOT02)
34	168	18902	BLADTUT07	1890243H1 (BLADTUT07), SARA01884F1, SATA00046F1, SARA03294F1, SARA02790F1

TABLE 1 (cont.)

Protein SEQ ID NO:	Nuc SEQ O:	Clone	Library	Fragments
35		19004	BLADTUT06	1900433H1 (BLADTUT06), SATA00396F1, SATA02742F1
36	170	19094	CONNTUT01	1909441H1 (CONNTUT01), 1398811F1 (BRAITUT08), 3039939H1 (BRSTNOT16), 3324740H1 (PTHYNOT03), 1442131F6 (THYRNOT03), 2254056H1 (OVARTUT01), 2199453T6 (SPLNFET02), 1692610F6 (COLNNOT23), 1698531H1 (BLADTUT05)
37	171	19322	COLNNOT16	1932226H1 (COLNNOT16), 2320569H1 (OVARNOT02), 1932226F6 (COLNNOT16), 2469455T6 (THP1NOT03), 2469455F6 (THP1NOT03), 1907140F6 (OVARNOT07), SATA02592F1
38	17	19326	COLNNOT16	1932647H1 (COLNNOT16), 1492745T1 (PROSNON01), 1492745H1 (PROSNON01), SASA02355F1, SASA00117F1, SASA00192F1
39	173	21242	BRSTNOT07	2124245H1 (BRSTNOT07), 1235393F1 (LUNGFET03), 1402264F6 (LATRTUT02), 1303990F1 (PLACNOT02), 1402264T6 (LATRTUT02)
40	74	21326	OVARNOT03	2132626H1 (OVARNOT03), 1723432T6 (BLADNOT06), 2132626R6 (OVARNOT03), 1736723T6 (COLNNOT22), 1504738F1 (BRAITUT07)
41	175	22806	PROSNON01	2280639H1 (PROSNON01), 1435330H1 (PANCNOT08), 1377560F6 (LUNGNOT10)
42		22923	BRAINON01	2292356H1 (BRAINON01), 4086827H1 (LIVRNOT06), 1754442F6 (LIVRTUT01), 3571126H1 (HEAPNOT01), 1601305F6 (BLADNOT03)
43	177	23493	COLSUCT01	2349310H1 (COLSUCT01), 2349310T6 (COLSUCT01)
44	17	23732	ADRENOT07	2373227H1 (ADRENOT07), 3316444H1 (PROSBPT03), 302685R6 (TESTNOT04), SASA02181F1, SASA01923F1, SASA03516F1
45	17	24576	ENDANOT01	2457682H1 (ENDANOT01), 2457682F6 (ENDANOT01)
46		24804	SMCANOT01	2480426H1 (SMCANOT01), 2480426F6 (SMCANOT01)

TABLE 1 (cont.)

Protein SEQ ID NO:	Nucl... SEQ ... O:	Clone I	Library	Fragments
47	1	25037	CONUTUT01	2503743H1 (CONUTUT01), 1853909H1 (HNT3AZT01), 1517619F1 (PANCTUT01), 1467896F6 (PANCTUT02), 490031F1 (HNT2AGT01), 1208654R1 (BRSTNOT02), 880544R1 (THYRNOT02)
48	1	25376	BONRTUT01	2537684H1 (BONRTUT01), 2005493H1 (TESTNOT03), 730969H1 (LUNGNOT03), 2537601F6 (BONRTUT01), 916487H1 (BRSTNOT04), 996135R1 (KIDNTUT01), 1920738R6 (BRSTTUT01), 1957710F6 (CONNNOT01)
49	1	25938	OVARTUT02	2593853H1 (OVARTUT02), 807497H1 (STOMNOT02), 914020R6 (STOMNOT02), 889992R1 (STOMTUT01)
50	1	26223	KERANOT02	2622354H1 (KERANOT02), 2623992H1 (KERANOT02), 1556510F6 (BLADTUT04)
51	185	264137	LUNGTUT08	2641377H1 (LUNGTUT08), 4341415H2 (BRAUNOT02), SBCA07049F3
52	1	26748	KIDNNOT19	2674857H1 (KIDNNOT19), 1872373H1 (LEUKNOT02), 470512R6 (MMLR1DT01), 1728547H1 (PROSNOT14), 3013651F6 (MUSCNOT07), SBCA01366F1, SBCA00694F1
53	1	27584	THP1AZS08	2758485H1 (THP1AZS08), 3097533H1 (CERVNOT03), 1578959F6 (DUODNOT01)
54	1	276329	BRSTNOT12	2763296H1 (BRSTNOT12), 3486025F6 (KIDNNOT31), SBDA07002F3
55	1	27794	OVARTUT03	2779436H1 (OVARTUT03), 2779436F6 (OVARTUT03), SBDA07009F3
56	190	28085	BLADTUT08	2808528H1 (BLADTUT08), 2611513F6 (THYMNOT04), SBDA07021T3
57	191	28092	BLADTUT08	2809230H1 (BLADTUT08), 2213849H1 (SINTFET03), 711706R6 (SYNORAT04), 958323R1 (KIDNNOT05), 030732F1 (THP1NOB01)
58	192	28168	BRSTNOT14	2816821H1 (BRSTNOT14), 3746964H1 (THYMNOT08), 2816821F6 (BRSTNOT14), 948722T6 (PANCNOT05), 807947R6 (STOMNOT02)

TABLE 1 (cont.)

Protein SEQ ID NO:	Nuc SEQ	le NO:	Clone	Library	Fragments
59			28172	BRSTNOT14	2817268H1 (BRSTNOT14), 3591308H1 (293TF5T01), 419522R1 (BRSTNOT01), 2073028F6 (ISLTNOT01), 1308781F6 (COLNFET02)
60			29231	SININOT04	2923165H1 (SININOT04), 2011630H1 (TESTNOT03), 1457250F1 (COLNFET02), 754668R1 (BRAITUT02), 1406510F6 (LATRTUT02)
61			29498	KIDNFET01	2949822H1 (KIDNFET01), SBDA07078F3
62			29921	KIDNFET02	2992192H1 (KIDNFET02), 2534324H2 (BRAINOT18), 2815255T6 (OVARNOT10), 1551107T6 (PROSNOT06), 1551107R6 (PROSNOT06)
63			29924	KIDNFET02	2992458H1 (KIDNFET02), 2618951H1 (GBLANOT01), 1479252F1 (CORPNOT02), 1879054H1 (LEUKNOT03), 1879054F6 (LEUKNOT03), 2215240H1 (SINTFET03), 1535968T1 (SPLNNOT04)
64	198		30447	HEAANOT01	3044710H1 (HEAANOT01), 3741773H1 (MENTNOT01), 859906X42C1 (BRAITUT03), 1534347F1 (SPLNNOT04), 1421122F1 (KIDNNOT09), 1303865F1 (PLACNOT02), 1704452F6 (DUODNOT02), 1251642F1 (LUNGFET03), 178169 4R6 (PGANNON02)
65			31204	LUNGTUT13	3120415H1 (LUNGTUT13), 1360123T1 (LUNGNOT12), 1375015H1 (LUNGNOT10)
66	200		12675	LUNGNOT01	126758H1 (LUNGNOT01), 126758X11 (LUNGNOT01), 811864T1 (LUNGNOT04)
67	201		67470	CRBLNOT01	674760H1 (CRBLNOT01), 3253976H1 (OVARTUN01), SAUA03387F1
68	202		12294	BRAITUT01	1229438H1 (BRAITUT01), 1230616H1 (BRAITUT01), 1461187R1 (PANCNOT04), 2493039H1 (ADRETUT05), 2891628H1 (LUNGFET04)
69	203		12369	LUNGFET03	1236935H1 (LUNGFET03), SBAA00983F1, SBAA02057F1, SBAA00170F1
70	20		13592	LUNGNOT12	1359283H1 (LUNGNOT12), SBAA01213F1, SBAA03934F1
71			14507	PENITUT01	551298F1 (BEPINOT01), 551298R1 (BEPINOT01), 1450703H1 (PENITUT01), 2748715H1 (LUNGTUT11)

TABLE 1 (cont.)

Protein SEQ ID NO:	Nucl[illegible]e SEQ I[illegible]:	Clone I	Library	Fragments
72	2[illegible]	19106[illegible]	CONNTUT01	1269346H1 (BRAINOT09), 1380872F1 (BRAITUT08), 1910668F6 (CONNTUT01), 1910668H1 (CONNTUT01), SATA02800F1, SATA03799F1, SARA02035F1
73	207	19551	CONNNOT01	1955143F6 (CONNNOT01), 1955143H1 (CONNNOT01)
74	[illegible]	19616[illegible]	BRSTNOT04	867025H1 (BRAITUT03), 1961637H1 (BRSTNOT04), 2809064T6 (BLADTUT08), 2938714H1 (THYMFET02), 2956402H1 (KIDNFET01), 3808735T6 (CONTTUT01)
75	209	19907[illegible]	CORPNOT02	1990762H1 (CORPNOT02), 1990762T3 (CORPNOT02), SBGA04911F1, SBGA01201F1, SBGA02205F1
76	210	19941	CORPNOT02	1994131H1 (CORPNOT02), 2645984F6 (OVARTUT04)
77	211	19977[illegible]	BRSTTUT03	1752307F6 (LIVRTUT01), 1853730H1 (HNT3AZT01), 1997745H1 (BRSTTUT03), SAZA00953F1
78	212	20090	TESTNOT03	2009035H1 (TESTNOT03), 2009035R6 (TESTNOT03)
79	213	20091	TESTNOT03	2009152H1 (TESTNOT03), 2009152R6 (TESTNOT03), 2783263H1 (BRSTNOT13)
80	[illegible]	20617	OVARNOT03	2061752H1 (OVARNOT03), 2061752T6 (OVARNOT03), 2732805H1 (OVARTUT04), SAZA01310F1, SAZA00830F1
81	[illegible]	20619	OVARNOT03	046580R1 (CORNNOT01), 746061R1 (BRAITUT01), 826996R1 (PROSNOT06), 2061933H1 (OVARNOT03)
82	216	20814	UTRSNOT08	2081422F6 (UTRSNOT08), 2081422H1 (UTRSNOT08), SBCA04793F1, SBCA05657F1, SBDA00065F1
83	217	21012	BRAITUT02	2101278H1 (BRAITUT02), SAXA00399F1, SAXA01284F1, SAXA01227F1
84	[illegible]	21213	BRSTNOT07	341437H1 (NEUTFMT01), 687136H1 (UTRSNOT02), 2121353H1 (BRSTNOT07), SASA01311F1

TABLE 1 (cont.)

Protein SEQ ID NO:	Nucl...e SEQ I.) NO:	Clone II	Library	Fragments
85	2	224173	PANCTUT02	833263H1 (PROSTUT04), 2241736H1 (PANCTUT02), SAZA01148F1, SASA03299F1, SASA01349F1
86	220	227193	PROSNON01	2271935H1 (PROSNON01), 2276774H1 (PROSNON01), 2760171T6 (THP1AZS08)
87	22	229534	BRSTNOT05	2295344H1 (BRSTNOT05), 3288561F6 (BONRFET01), SBGA01801F1
88	22	230399	BRSTNOT05	905482T1 (COLNNOT08), 1858636F6 (PROSNOT18), 2303994H1 (BRSTNOT05)
89	2	249780	ADRETUT05	2497805F6 (ADRETUT05), 2497805H1 (ADRETUT05)
90	2	264636	LUNGTUT11	1754702H1 (LIVRTUT01), 2640776T6 (LUNGTUT08), 2646362H1 (LUNGTUT11), 3356773H1 (PROSTUT16)
91	2	265714	LUNGTUT09	2657146F6 (LUNGTUT09), 2657146H1 (LUNGTUT09)
92	2	275578	THP1AZS08	288436R1 (EOSIHET02), 1252824F6 (LUNGFET03), 1305549H1 (PLACNOT02), 1364975R1 (SCORNON02), 2018293H1 (THP1NOT01), 2047320H1 (THP1T7T01), 2184537F6 (SININOT01), 2755786H1 (THP1AZS08), 4111022H1 (PROSBPT07)
93	22	283124	TLYMNOT03	2831245H1 (TLYMNOT03), SBMA01396F1
94	228	311625	LUNGTUT13	126263F1 (LUNGNOT01), 2729942H1 (OVARTUT04), 3116250H1 (LUNGTUT13)
95	229	312963	LUNGTUT12	3129630F6 (LUNGTUT12), 3129630H1 (LUNGTUT12), SBDA06436F1
96	2	00763	HMC1NOT01	007632H1 (HMC1NOT01), 007632R6 (HMC1NOT01), 007632T6 (HMC1NOT01)
97	231	123696	LUNGFET03	1236968H1 (LUNGFET03), SBAA02713F1, SBAA03203F1, SBAA04196F1
98	232	133415	COLNNOT13	776410R1 (COLNNOT05), 1334153H1 (COLNNOT13), 1334153T1 (COLNNOT13), 1800085F6 (COLNNOT27), 2701948H1 (OVARTUT10)

TABLE 1 (cont.)

Protein SEQ ID NO:	Nucl ide SEQ ID NO:	Clone	Library	Fragments
99	233	13969	BRAITUT08	864113H1 (BRAITUT03), 876139R1 (LUNGAST01), 1268313F1 (BRAINOT09), 1351348T1 (LATRTUT02), 1396975H1 (BRAITUT08), 1485768F6 (CORPNOT02), 1815364F6 (PROSNOT20)
100	234	150174	SINTBST01	079080R1 (SYNORAB01), 1501749H1 (SINTBST01), 1724970H1 (PROSNOT14)
101	235	15752	LNODNOT03	081858R1 (SYNORAB01), 1575240H1 (LNODNOT03), 3451462R6 (UTRSNON03)
102	236	16478	PROSTUT09	1647884H1 (PROSTUT09), 1647884T6 (PROSTUT09), 3998922R6 (HNT2AZS07)
103	237	16611	BRSTNOT09	720941X17 (SYNOOAT01), 1661144H1 (BRSTNOT09), 2181782H1 (SININOT01)
104	238	16854	PROSNOT15	755203R1 (BRAITUT02), 1226185T1 (COLNNOT01), 1300837F1 (BRSTNOT07), 1685409H1 (PROSNOT15), 1705256H1 (DUODNOT02)
105	239	17314	BRSTTUT08	1731419H1 (BRSTTUT08), 1731419X319T3 (BRSTTUT08), 1731419X322F1 (BRSTTUT08), 1731419X326F1 (BRSTTUT08), 1731419X329F1 (BRSTTUT08), 1733786F6 (BRSTTUT08), SZAH01494F1
106	240	26502	BRSTNOT14	1680316T6 (STOMFET01), 2650265H1 (BRSTNOT14), 2650265T6 (BRSTNOT14), 2760588R6 (BRAINOS12)
107	241	26771	KIDNNOT19	1592129H1 (CARGNOT01), 2645962H1 (OVARTUT04), 2677129F6 (KIDNNOT19), 2677129H1 (KIDNNOT19), 2910973H1 (KIDNTUT15), 4571722H1 (PROSTMT02), 4906791H2 (TLYMNOT08)
108	242	31510	ADRENON04	3150857T6 (ADRENON04), 3151073H1 (ADRENON04), 3151073R6 (ADRENON04)
109	243	31700	BRSTNOT18	3170095F6 (BRSTNOT18), 3170095H1 (BRSTNOT18)

TABLE 1 (cont.)

Protein SEQ ID NO:	Nucl SEQ I	e O:	Clone I	Library	Fragments
110	24		347516	LUNGNOT27	079680F1 (SYNORAB01), 443811T6 (MPHGNOT03), 1509356T6 (LUNGNOT14), 1873596F6 (LEUKNOT02), 2440867H1 (EOSITXT01), 3475168H1 (LUNGNOT27)
111	2		383689	DENDTNT01	446637H1 (MPHGNOT03), 1219376R6 (NEUTGMT01), 3735467F6 (SMCCNOS01), 3735467T6 (SMCCNOS01), 3836893H1 (DENDTNT01)
112	2		407215	KIDNNOT26	2129415T6 (KIDNNOT05), 4072159F6 (KIDNNOT26), 4072159H1 (KIDNNOT26)
113	24		100391	BRSTNOT03	620937R6 (PGANNOT01), 1003916H1 and 1003916R6 (BRSTNOT03), 1413623H1 (BRAINOT12), 1435945F1 (PANCNOT08), 1479127F1 (CORPNOT02), 1969146R6 (BRSTNOT04), 2517587F6 (BRAITUT21), 2967848H1 (SCORNOT04)
114	2		209349	PANCNOT04	489651H1 (HNT2AGT01), 1265353T1 (SYNORAT05), 1431505R6 (BEPINON01), 1605237F6 (LUNGNOT15), 2093492H1 and 2093492T6 (PANCNOT04), 4195560H1 (COLITUT02)
115	249		210878	BRAITUT03	2108789H1 and 2108789R6 (BRAITUT03), 2182008T6 (SININOT01), 3255751R6 and 3255751T6 (OVARTUN01)
116	2		217140	ENDCNOT03	037241F1 (HUVENOB01), 1821492F6 (GBLATUT01), 2055814T6 (BEPINOT01), 2171401F6 and 2171401H1 (ENDCNOT03), 2668952F6 (ESOGTUT02), 3140313H1 and 3140313T6 (SMCCNOT02), 5031775H1 (EPIBTXT01)
117	2		221252	SINTFET03	187596R6 and 187596T6 (CARDNOT01), 919634R6 (RATRNOT02), 1992331H1 (CORPNOT02), 2062034H1 (OVARNOT03), 2212530F6 and 2212530H1 (SINTFET03), 2520479H1 (BRAITUT21), 2878284F6 (THYRNOT10), 2992354H1 (KIDNFET02), 4020719F6 (BRAXNOT02)
118	2		225303	OVARTUT01	2253036H1 and 2253036R6 (OVARTUT01)

TABLE 1 (cont.)

Protein SEQ ID NO:	Nuc' SEQ	le NO:	Clone '	Library	Fragments
119			22801	PROSNON01	482326H1 (HNT2RAT01), 934345H1 (CERVNOT01), 1379358F1 and 1379358T1 (LUNGNOT10), 1438562T1 (PANCNOT08), 1467511F6 (PANCTUT02), 1568138F1 (UTRSNOT05), 1636106T6 (UTRSNOT06), 2134534F6 (ENDCNOT01), 2280161H1 and 2280161X19F1 (PROSNON01), 2789845F6 (COLNTUT16), 3096938H1 (CERVNOT03), 3774621F6 (BRSTNOT25), 4222971H1 (PANCNOT07), 5111983H1 (ENDITXT01), 5324177H1 (FIBPFEN06)
120			22874	BRAINON01	1454588F1 (PENITUT01), 1593332F6 (BRAINOT14), 2287485H1 and 2287485R6 (BRAINON01), 3765992H1 (BRSTNOT24), 4374293H1 (CONFNOT03), 4937931H1 (PROSTUS18), SBCA01722F1
121			23803	ISLTNOT01	2380344F6 and 2380344H1 (ISLTNOT01), 2888536T3 (LUNGFET04), SASA03644F1, SASA03689F1
122			23831	ISLTNOT01	956296R1 (KIDNNOT05), 1342250F1 (COLNTUT03), 1468046F1 and 1468046T1 (PANCTUT02), 2383171H1 (ISLTNOT01), SBYA05452U1, SBYA01369U1
123			23960	THP1AZT01	2396046F6, 2396046H1 and 2396118T6 (THP1AZT01)
124			24565	ENDANOT01	2456587H1 and 2456587T6 (ENDANOT01), 2872569H1 (THYRNOT10), SBCA03778F1, SBDA00115F1, SBCA02401F1, SBCA03351F1, SBCA05164F1, SBCA04783F1, SBCA00155F1, SBCA04141F1
125			24848	BONRTUT01	1234970T1 (LUNGFET03), 1338090F6 (COLNNOT13), 2484813H1 (BONRTUT01), SBCA00053F1, SBCA02064F1, SBCA02151F1, SBCA03770F1, SBCA04866F1, SBCA03406F1
126	260		24938	ADRETUT05	2493851H1 (ADRETUT05), 3805916F6 (BLADTUT03), 4500439H1 and 4500748H1 (BRAVTXT02), 5120601H1 (SMCBUNT01)
127			24957	ADRETUT05	603447R1 (BRSTTUT01), 2495719H1 (ADRETUT05), 2917493F6 (THYMFET03), 4647103H1 (PROSTUT20), SBRA04984D1

TABLE 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
128	2	26141	GBLANOT01	1833135R6 (BRAINON01), 1966515R6 (BRSTNOT04), 2331103R6 (COLNNOT11), 2614153H1 (GBLANOT01), 2656691F6 (LUNGTUT09), 3951176H1 (DRGCNOT01)
129	26	265518	THYMNOT04	2655184H1 (THYMNOT04), SBDA05215F1, SBDA05213F1, SBDA01516F1
130	2	284836	BRSTTUT13	1297974F1 and 1297974T6 (BRSTNOT07), 2630138F6 (COLNTUT15), 2848362H1 (BRSTTUT13)
131	2	284990	BRSTTUT13	1541617R1 and 1541617T1 (SINTTUT01), 2684504F6 and 2684504T6 (LUNGNOT23), 2796805H1 (NPOLNOT01), 2849906H1 (BRSTTUT13)
132	2	289913	DRGCNOT01	2899137H1 (DRGCNOT01), 3026490F6 and 3026490T6 (HEARFET02), 3483359H1 (KIDNNOT31)
133	2	298622	CARGDIT01	1740227T6 (HIPONON01), 2986229H1 (CARGDIT01)
134	2	322208	COLNNON03	1754079F6 (LIVRTUT01), 3222081H1 (COLNNON03), 4053813T6 (SPLNNOT13), 4230282H1 (BRAMDIT01), SBDA07029F3

TABLE 2

Protein SEQ ID NO:	Ami... Aci... Resid...	Potential Phosphorylation ...	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
1	88	T83 S38 T76		M1 - A21		Signal Peptide HMM
2	12...	S30 S...0 T47 T1 W125		M1 - F28		Signal Peptide HMM
3	11...	T70		M1 - T18		Signal Peptide HMM
4	11...	S32 T...4	N58	M1 - A29		Signal Peptide HMM
5	78	T27 S39 S39 S... ...2 T27 ...28 S57		M1 - R24		Signal Peptide HMM
6	88	T55 S...0 S40 T...	N34	M1 - N21		Signal Peptide HMM
7	22...	S220 ...70 S83 T... S134 ...141 T15... Y123	N100	M1 - Q20		Signal Peptide HMM
8	198	S62 T...123 S142 S62 ...100 Y85	N60	M1 - A28		Signal Peptide HMM
9	65...	T48		M1 - A29		Signal Peptide HMM
10	154			M1 - A29		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Ami Aci Resid s	Potential Phosp horylatior	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
11	23	T116 26 T79 T T182 188 T19 T206 60 S123 S213	N128	M1 - A19		Signal Peptide HMM
12	225	T158 128	N166	M1 - G27		Signal Peptide HMM
13	11	S41		M1 - A23		Signal Peptide HMM
14	253	S49 T63 S92 T1 S127 T239	N42 N47 N72 N207	M1 - T20		Signal Peptide HMM
15	171	S43 S 4 T114		M88 - R112		Signal Peptide HMM
16	79	S38 S 3	N37	M1 - G19		Signal Peptide HMM
17	71	T64 T 7		M1 - C19		Signal Peptide HMM
18	189	S36 8 T133	N121 N171	M1 - A21		Signal Peptide HMM
19	80	S76		M1 - C19		Signal Peptide HMM
20	80			M1 - G25		Signal Peptide HMM
21	84	S39 S 3 S60		M1 - G21		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
22	171	S41 T150		M3 - A21		Signal Peptide HMM
23	243	S3 S4[illegible] T75 S86 S223 S36 S92 S Y40 Y110	N97	M1 - C25		Signal Peptide HMM
24	311	T5 S76 T82 T9[illegible] T109 S121 T13[illegible] T17[illegible] S184 S11 S75 S[illegible]4 T132 S S274 Y69	N49 N91 N108 N128 N135 N190	M1 - A32		Signal Peptide HMM
25	57			M1 - L29		Signal Peptide HMM
26	82	S46 Y26		M1 - S18		Signal Peptide HMM
27	11[illegible]			M1 - G34		Signal Peptide HMM
28	327	S93 S[illegible]0 S167 S S89 T105 T214 S[illegible] T318	N138 N206	M1 - E25		Signal Peptide HMM
29	13[illegible]	S63	N105	M1 - E29		Signal Peptide HMM
30	12	S21 S[illegible]5 T93		M1 - G20		Signal Peptide HMM

TABLE 2

Protein SEQ ID NO:	Amino Ac Resi	Potential Phosphorylatio	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
1	8	T83 38 T76		M1 - A21		Signal Peptide HMM
2	12	S30 40 T47 T W125		M1 - F28		Signal Peptide HMM
3	111	T70		M1 - T18		Signal Peptide HMM
4	11	S32 T64	N58	M1 - A29		Signal Peptide HMM
5	7	T27 S39 S39 S T27 S28 S57		M1 - R24		Signal Peptide HMM
6	88	T55 S30 S40 T	N34	M1 - N21		Signal Peptide HMM
7	227	S220 370 S83 S134 S141 T15 Y123	N100	M1 - Q20		Signal Peptide HMM
8	19	S62 T123 S142 S62 T100 Y85	N60	M1 - A28		Signal Peptide HMM
9	65	T48		M1 - A29		Signal Peptide HMM
10	154			M1 - A29		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Ami Aci Resid	Potential Phosp orylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
11	23	T116 T26 T79 T T182 T188 T19 T206 S60 S123 S213	N128	M1 - A19		Signal Peptide HMM
12	225	T158 S128	N166	M1 - G27		Signal Peptide HMM
13	11	S41		M1 - A23		Signal Peptide HMM
14	253	S49 T 3 S92 T1 S127 T239	N42 N47 N72 N207	M1 - T20		Signal Peptide HMM
15	171	S43 S94 T114		M88 - R112		Signal Peptide HMM
16	78	S38 S43	N37	M1 - G19		Signal Peptide HMM
17	71	T64 T 7		M1 - C19		Signal Peptide HMM
18	188	S36 T53 T133	N121 N171	M1 - A21		Signal Peptide HMM
19	80	S76		M1 - C19		Signal Peptide HMM
20	80			M1 - G25		Signal Peptide HMM
21	84	S39 S53 S60		M1 - G21		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
22	171	S41 T150		M3 - A21		Signal Peptide HMM
23	243	S3 T75 S86 S2 36 S92 S2 Y4 110	N97	M1 - C25		Signal Peptide HMM
24	311	T5 T82 T93 T109 121 T137 T170 S184 S11 S75 S 4 T132 S S274 69	N49 N91 N108 N128 N135 N190	M1 - A32		Signal Peptide HMM
25	57			M1 - L29		Signal Peptide HMM
26	82	S46 6		M1 - S18		Signal Peptide HMM
27	115			M1 - G34		Signal Peptide HMM
28	327	S93 S50 S167 S S89 T105 T214 T318	N138 N206	M1 - E25		Signal Peptide HMM
29	133	S63	105	M1 - E29		Signal Peptide HMM
30	129	S21 S 5 T93		M1 - G20		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Amin Ac Resid	Potential Phosphorylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
31	472	S164 T32 S42 T T154 155 T23 T262 271 T33 T376 402 S42 S435 441 S19 T327 378	N61 N179 N353 N356 N396	M1 - G20	hematopoietic lineage switch 2 (g3169729)	Signal Peptide HMM BLAST - GenBank
32	93	T21		M1 - A18		Signal Peptide HMM
33	92	S57 S		M1 - G47		SPScan
34	142	T6 T1 T135		M9 - G40		Signal Peptide HMM
35	89	T15 S 8 S66		M1 - A19		Signal Peptide HMM
36	56	T7 T S150 T S228 257 S35 S474 529 S53 T186 219 S36 Y523	N163 N184 N379	M1 - E34		SPScan
37	197	T80 S163		M1 - G28		Signal Peptide HMM
38	437	T47 T146 S233 S403 43 S130 S339 364	N46 N189 N382	M1 - A21		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
39	330	S197 T[illegible]9 T150 S[illegible] T214 T215 T49 S[illegible] S237	N46 N64 N166 N191	M1 - G28		Signal Peptide HMM
40	148	T73 S1[illegible]1	N29 N58 N71 N103	M1 - R24	receptor-activity-modifying protein (RAMP; g4165368)	Signal Peptide HMM BLAST - GenBank
41	188	S49		M1 - V25		Signal Peptide HMM
42	222	S89 S1[illegible]5 T174 T[illegible] T83 S1[illegible]5		M1 - S24		Signal Peptide HMM
43	111	S54 S2[illegible] S98 S5[illegible] T104		M1 - T23		Signal Peptide HMM
44	341	T29 S1[illegible]6 T120 S[illegible] S195 S[illegible]7 S47 T5[illegible] S136 S[illegible]23 S230 S281		M1 - G22		Signal Peptide HMM
45	148	S21 T6[illegible] T63 A1	N40	M1 - G23		Signal Peptide HMM
46	87	S65		M1 - P18		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residue	Potential Phosphorylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
47	383	T77 S95 S108 S2 S351 S121 S124 S153 T187	N93 N207	M1 - P23		Signal Peptide HMM
48	109	S25 S2		M1 - L18		Signal Peptide HMM
49	185	S62		M1 - A20		Signal Peptide HMM
50	110	T100 T73 S97 Y	N71	M1 - C21		Signal Peptide HMM
51	126	S17 S110		M1 - G18		Signal Peptide HMM
52	489	S205 T31 S86 T S7 T447	N250 N321 N463	M1 - L25	putative involvement in cell wall structure or biosynthesis (g3738170)	Signal Peptide HMM BLAST - GenBank
53	197	T55 S31 S46 S6 S108 T119 T167 S194 S2 S34 T1		M1 - A26		Signal Peptide HMM
54	84	S65 S T41 S5 S81	N39	M1 - G25		Signal Peptide HMM
55	97	S56		M1 - A22		Signal Peptide HMM
56	140	S29		M1 - P23		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Am... A... Residu...	Potentia... Phos... horylatio...	...s	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
57	28...	S53 S108 T216 S277	...53	N153	M1 - A25		Signal Peptide HMM
58	26...	S62 T166 S62 Y246		N190	M1 - G28	3-acylating enzyme (Q44449)	Signal Peptide HMM BLAST - GENESEQ
59	1...	S120 T154 T3... S174	...		M1 - C22		Signal Peptide HMM
60	2...	S98 ... 136 T67 S234 ...237			M55 - E84ß		SPScan
61	...	T68		N67	M1 - G18		Signal Peptide HMM
62	2...	T21 S117 S12...			M1 - G27		Signal Peptide HMM
63	450	S107 S97 S146 S440 S245 T3... S304 S399			M1 - G18		Signal Peptide HMM
64	322	T14... T214 T1... S35 ...15 T145 ... S297 ...300 T3		N53 N130 N289	M1 - G23		Signal Peptide HMM
65	1...	S38 ...25 S75			M1 - A18		Signal Peptide HMM

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residue	Potential Phosphorylation S	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
66	93			M1 through about S18 Transmembrane: M1 through about Y17		SPscan HMM
67	71	S23 S64		M1 through about A24		SPscan HMM MOTIFS
68	394	S392 S393 S31 S S179 S334 T338 S358 T383 Y323	N53	M1 through about S31 Transmembrane: about M159 through about F178 about F109 through about S127 about F225 through about V243		SPScan HMM MOTIFS
69	72	S59	N69	M1 through about S23 Transmembrane: M1 through about L16		SPscan HMM MOTIFS
70	71	S11 T26		M1 through about Q18		SPscan HMM MOTIFS
71	247	S41 T79		M1 through about S25		SPscan HMM MOTIFS
72	73	S56		M1 through about G27		SPscan HMM MOTIFS

TABLE 2 (cont.)

Protein SEQ ID NO:	Amin Aci. Residue	Potential Phosphorylation S	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
73	70			M1 through about G20		SPscan HMM
74	67			M1 through about G30		SPscan HMM
75	91			M1 through about G26		SPScan
76	56	T29 S[illegible] T51		M1 through about S19		SPscan HMM MOTIFS
77	112	S62 S6[illegible]		M1 through about G27 Transmembrane: about W79 through about H97		SPscan HMM MOTIFS
78	54		N48	M1 through about N34		SPscan HMM MOTIFS
79	57	T33 R[illegible]		M1 through about C18		SPscan HMM MOTIFS
80	52			M1 through about S30		SPscan HMM MOTIFS

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residue	Potential Phosphorylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
81	64	T43 Y27		M1 through about S41		SPscan HMM MOTIFS
82	65	S45		M1 through about A31 Transmembrane: about L38 through about F55		SPscan HMM MOTIFS
83	56			M1 through about E23		SPscan HMM
84	120	S69 S1[illegible]9	N89 N95	M1 through about A38 Transmembrane: about L23 through about T41		SPscan HMM MOTIFS
85	67	S28		M1 through about K30 Microbodies C-terminal targetting signal: A65KV		SPscan HMM MOTIFS
86	62	S29 S[illegible] S46	[illegible]40	M1 through about S29		SPscan HMM MOTIFS
87	75	S25 S[illegible]		M1 through about L19 Transmembrane: about I3 through about G20		SPscan HMM MOTIFS

TABLE 2 (cont.)

Protein SEQ ID NO:	Amin. Aci' Resid	Potential Phosp' orylatior	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
88	8(	T28		M1 through about A20		SPscan HMM MOTIFS
89	5(	S11		M1 through about C48		SPscan HMM MOTIFS
90	116	S38		M1 through about G22		SPscan HMM MOTIFS
91	67	S43		M1 through about P21		SPscan HMM MOTIFS
92	53.	S415 S52 T77 S T178 T228 S28 S320 S332 S38 T401 T424 S48 S207 S230 S35 T410 Y263 Y36	N226	M1 through about S18 Tyrosine specific protein phosphatases signature: about V328 through about F340		SPScan BLOCKS PRINTS MOTIFS
93	58			M1 through about S25		SPscan HMM
94	11(	S39		M1 through about S22 Transmembrane: about V3 through about S21		SPscan HMM MOTIFS

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residue	Potential Phosphorylation S	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
95	128	S91		M1 through about G31 Transmembrane: about F108 through about L126		SPscan HMM MOTIFS
96	124	T115 T43 S91		M1-S20 P116-V124 (urotensin II signature)		SPScan HMM Motifs BLOCKS BLAST
97	182	S28 T70 S172 S2 S32 S48 S108 S1		M1-S23, M1-S25		SPScan HMM Motifs
98	237	S55 S8 S121 S1	N45 N73 N107 N118 N132 N172 N175 N185	M1-A16, M1-S21 C40-C198 (cysteine spacing pattern similar to that of RoBo-1)		SPScan HMM Motifs BLAST
99	160	S36 S5 T143		M1-A27		SPScan HMM Motifs
100	148	T76 S6 Y103		M1-S30, M1-G31		SPScan HMM Motifs
101	170	S78 T4 T30 S130 S29 T2		M1-A23, M1-L28		SPScan HMM Motifs

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
102	150	S50 S73 S91		M1-A26, M1-S28		SPScan HMM Motifs
103	142	T57 T8[illegible]		M1-A25, M1-G26		SPScan HMM Motifs
104	110	T3		M1-G18, M1-T25		SPScan HMM Motifs
105	120	T29 S40 S72		M1-G22, M1-A20		SPScan HMM Motifs
106	135	T115 S38 T41	N32 N101	M1-G26, M1-C25		SPScan HMM Motifs
107	301	S53 S217 S240 S T224		M1-A22		SPScan HMM Motifs
108	103	S88 T93 S84		M1-P19, M1-L22		SPScan HMM Motifs
109	95	T82 S52 S77	N50	M1-T15, M1-P19		SPScan HMM Motifs

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residue	Potential Phosphorylation S	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
110	113	T84 S4		M1-P19, M1-A24		SPScan HMM Motifs
111	234	S179 S184 S51 T T158 S168 T228	N146 N191 N194	M1-A20	NK cell activating receptor (g4493702)	SPScan HMM Motifs BLAST - GenBank
112	119	S39 T61		M1-G30, M1-G27		SPScan HMM Motifs
113	200	S51 T4 S191		M1-G26 Signal Peptide	Signal Peptide Containing Protein, Homology with KIAA0206	SPScan Motifs BLAST
114	225			M1-Q29 Signal Peptide	Signal Peptide Containing Protein	SPScan
115	155	S29		M1-A20 Signal Peptide	Signal Peptide Containing Protein	HMM Motifs
116	468	S143 T 56 T227 S235 T 71 T293 T436 S 53 S117 148 213 S263 417 73	N280 N384	M1-G23 Signal Peptide	Signal Peptide Containing Protein	SPScan Motifs
117	403	S19 0 S69 S1 T171 T 7 S393 Y Y378	387	M1-A24 Signal Peptide	Signal Peptide Containing Protein	HMM Motifs

TABLE 2 (cont.)

Protein SEQ ID NO:	Ami Aci Resid	Potential Phosp orylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
118	131	T131 S24 T79 T T123 T127	N116	M1-G25 Signal Peptide	Signal Peptide Containing Protein	SPScan Motifs
119	556	T176 S192 S196 T220 S344 S36 S476 T501 S529 S541 T548 T55 S115 121 T38 T424 500 Y104	N62 N79 N127 N157 N160	M1-P21 Signal Peptide L226-W244, Y402-W422, V375-L392 and Y355-I376 Transmembrane Domains	Signal Peptide Containing Protein, Weakly similar to Putative Transmembrane Protein (PTM1) Precursor	SPScan Motifs HMM BLAST
120	51	T457 T80 S86 T T372 T420 S44 T102 S112 T24 S297 S353 S47	N100 N168 N319	M1-G24 Signal Peptide	Signal Peptide Containing Protein,	SPScan Motifs
121	109	T46 S 3 T12		M1-S15 Signal Peptide	Signal Peptide Containing Protein	SPScan MotifS
122	43	S57 T. 20 S339 S100 239		M1-L25 Signal Peptide	Signal Peptide Containing Protein, Weakly similar to OXA1L	SPScan Motifs BLAST
123	14			M1-W16 Signal Peptide	Signal Peptide Containing Protein	SPScan

TABLE 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residu	Potential Phosphorylation	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
124	643	T8 S28 S77 T16 T199 T 35 S252 T320 S 02 T413 S414 S 58 S22 T S56 S62 S120 T1 S329 T 23 S475 S574 Y 26	N251	M1-S28 Signal Peptide, D37-C81, W380-C437, W440-C492 and F526-C583 Thrombospondin Type 1 Domains	Signal Peptide Containing Protein, Thrombospondin Type 1 Protein	SPScan Motifs Pfam BLAST
125	568	S510 T 24 T80 S T153 T165 S232 S248 S 62 T300 T334 S 80 S446 T19 T6 S127 S T436 T 31 S554 T564 Y135 Y48	N322	M1-T19 Signal Peptide	Signal Peptide Containing Protein	SPScan Motifs
126	125	T62 S 7 T36		M1-R32 Signal Peptide, V4-L53 Glycosyl Hydrolase Family 9 Active Site Signature	Signal Peptide Containing Protein, Glycosyl Hydrolase Protein	SPScan Motifs PROFILE-SCAN
127	196	T105 T 7 T56 S		M1-S26 Signal Peptide, H79-H123 Ribosomal Protein S18 Signature	Signal Peptide Containing Protein, Ribosomal Protein S18	SPScan Motifs BLAST Pfam PROFILE-SCAN
128	214	S112 S 31	N37 N92	M1-S35 Signal Peptide	Signal Peptide Containing Protein, Homology with GTP Binding Protein	SPScan Motifs BLAST

TABLE 2 (cont.)

Protein SEQ ID NO:	Amin Acid Residu	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
129	88			M1-S24 Signal Peptide	Signal Peptide Containing Protein	HMM
130	260	S146 S179 S192 S239 S270 T126	N50 N109	M1-A48 Signal Peptide, G59-S142 Immunoglobulin Domain	Signal Peptide Containing Protein, Immunoglobulin Superfamily Protein	SPScan Motifs Pfam
131	295	T176 T56 S72 S1 S256 S 7		M1-A30 Signal Peptide	Signal Peptide Containing Protein	SPScan Motifs
132	183	S11 T41 T42 S8		M1-W24 Signal Peptide, E131-K168 and C105-H115 Adrenodoxin Iron-Sulfur Binding Signature, C111-V116 Cytochrome C Heme Binding Signature, N69-A162 Iron-Sulfur Cluster Binding Domain	Signal Peptide Containing Protein, Adrenodoxin Family Iron-Sulfur Binding Protein, and Cytochrome C Family Heme Binding Protein	HMM Motifs BLOCKS PRINTS Pfam
133	113	S93 T8 Y9		M1-G30 Signal Peptide, V28-L74 PF00646 F-Box Domain	Signal Peptide Containing Protein, PF00646 F-Box Protein	SPScan Motifs Pfam
134	160	T46 T55 S65 S1 T125 T16		M1-A27 Signal Peptide	Signal Peptide Containing Protein, F45G2.10 and Yhr122wp Homology	SPScan Motifs BLAST

TABLE 3

Nucleotide SEQ ID NO:	Tissue Expression (Fraction Total)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
135	Hemat ietic/Immune (1.0	Inflammation (1.000)	pBLUESCRIPT
136	Hem ietic/Immune (0.7 rdiovascular (0.250)	Inflammation (0.750) Cancer (0.250)	pSPORT1
137	Nerv 1.000)	Trauma (1.000)	pSPORT1
138	Musc keletal (1.000)	Inflammation (1.000)	pSPORT1
139	Gast stinal (0.714) Card scular (0.143) Repro tive (0.143)	Cancer (0.714) Trauma (0.143)	pSPORT1
140	Nerv (1.000)	Neurological (0.500) Trauma (0.500)	pSPORT1
141	Repr tive (0.293) Gastro inal (0.146) Hem ietic/Immune (0.1	Cancer (0.524) Inflammation (0.256) Fetal (0.146)	pSPORT1
142	Repr tive (0.266) Gastro tinal (0.170) Nerv 0.138)	Cancer (0.479) Inflammation (0.277) Fetal (0.181)	pINCY
143	Repr tive (0.417) Nervo 292) Dev ental (0.125)	Cancer (0.417) Inflammation (0.250) Fetal (0.167)	pINCY
144	Repr tive (0.321) Cardio ular (0.143) Deve pmental (0.143)	Cancer (0.464) Fetal (0.214) Inflammation (0.143)	pINCY
145	Repr tive (0.600) Gastro tinal (0.400)	Cancer (0.400) Trauma (0.400) Inflammation (0.200)	pINCY
146	Card cular (0.400) Derm ic (0.200) Ner 0.200)	Cancer (0.600) Fetal (0.600)	pINCY
147	Dev ental (0.667) Gast inal (0.333)	Fetal (0.667) Cancer (0.333)	pINCY
148	Repr tive (0.256) Nervo) Card scular (0.137)	Cancer (0.479) Inflammation (0.214) Fetal (0.145)	pINCY
149	Repr tive (0.244) Nervo) Hem ietic/Immune (0.1	Cancer (0.433) Inflammation (0.322) Fetal (0.156)	pINCY

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissue ...pression (Fraction ...)	...)	Disease/Condition-Specific Expression (Total of Fraction)		Vector
150	Cardi... ...ular (0.923) Deve...	...tal (0.077)	Cancer (0.692) Fetal (0.154) Inflammation (0.154)		pINCY
151	Repro... ...ive (0.215) Nervou... Gastro... ...stinal (0.177)	...00)	Cancer (0.494) Inflammation (0.278) Trauma (0.152)		pINCY
152	Reprod... ...ive (0.200) Nervou... Hema... ...ietic/Immune (0.14	...71)	Inflammation (0.371) Cancer (0.229) Fetal (0.200)		pINCY
153	Repro... ...ive (0.333) Nervou... Hema... ...ietic/Immune (0.13	...57)	Cancer (0.549) Inflammation (0.176) Fetal (0.137)		pINCY
154	Gastro... ...tinal (0.500) Urol...	...0.167)	Inflammation (0.667) Cancer (0.167) Trauma (0.167)		pINCY
155	Gastr... ...stinal (0.429) Repr... Nervou... ...0.143)	...ve (0.286)	Inflammation (0.429) Cancer (0.286) Trauma (0.143)		pINCY
156	Repro... ...ive (1.000)		Cancer (0.500) Inflammation (0.500)		pINCY
157	Hema... ...poietic/Imm... ...ne (0.3... Gastro... ...stinal (0.115)	...roductive (0.154)	Cancer (0.404) Inflammation (0.404) Fetal (0.212)		pINCY
158	Repro... ...ive (0.23...) Hemat... Gastr... ...stinal (0.132)	...c/Immune (0.217)	Cancer (0.415) Inflammation (0.358) Fetal (0.142)		pINCY
159	Gastr... ...stinal (1.000)		Cancer (1.000)		pSPORT1
160	Develop... ...ental (0.500) Hema... Nervou... ...0.250)	...tic/Immune (0.250)	Fetal (0.500) Inflammation (0.250) Trauma (0.250)		pINCY
161	Hematopoietic/Imm...une (0.25... Nervous (0.208)	...productive (0.250)	Cancer (0.583) Fetal (0.292) Inflammation (0.250)		pINCY
162	Gastr... ...stinal (0.412) Repr... Cardi... ...cular (0.088)	...ive (0.412)	Cancer (0.735) Inflammation (0.176) Fetal (0.029)		pINCY

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissue ression (Fraction o' ol)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
163	Repro' e (0.29') Cardiov r (0.170) Nervous (0.149	Cancer (0.532) Inflammation (0.213) Fetal (0.191)	pINCY
164	Gastro tinal (0.333) Hem tic/Immune (0.333) Repro ve (0.333)	Cancer (0.667) Inflammation (0.333)	pINCY
165	Reprod ive (0.29) Gastroi inal (0.159) Nervo .148)	Cancer (0.534) Inflammation (0.284) Fetal (0.091)	pINCY
166	Hemat etic/Imm une (0.53 diovascular (0.077) Reproductive (0.07)	Inflammation (0.731) Cancer (0.154) Fetal (0.154)	pINCY
167	Reproductive (0.48) Gastroi nal (0.121) Nervous (0.103)	Cancer (0.672) Inflammation (0.155)	pINCY
168	Gastroin tinal (0.222) Hem etic/Immune (0.222) Nervous (.148)	Cancer (0.519) Inflammation (0.370) Fetal (0.259)	pINCY
169	Urolo .000)	Cancer (0.333) Fetal (0.333) Inflammation (0.333)	pINCY
170	Repro ve (0.21) Gastroi al (0.179) Nervo .143)	Cancer (0.643) Inflammation (0.143) Fetal (0.107)	pINCY
171	Repro ve (0.26) Develo al (0.174) Nerv 174)	Cancer (0.391) Fetal (0.304) Inflammation (0.217)	pINCY
172	Repro ve (0.35) Gastroi nal (0.321) Cardiov ular (0.071)	Cancer (0.571) Inflammation (0.286) Fetal (0.107)	pINCY
173	Repr ve (0.30) Nervou 1) Cardiov ular (0.1 9)	Cancer (0.387) Inflammation (0.323) Fetal (0.226)	pINCY
174	Repro uct ve (0.22) Nervou 8) Card ar (0. 7)	Cancer (0.521) Inflammation (0.312) Trauma (0.146)	pSPORT1

TABLE 3 (cont.)

Nucleotide SEQ ID NO	Tissue Expression (Fraction o	l)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
175	Repr ive (0.44) Develo Cardi cular (0.111)	al (0.167)	Cancer (0.556) Fetal (0.278) Trauma (0.111)	pSPORT1
176	Repr ive (0.29) Gastroi Cardi cular (0.118)	al (0.176)	Cancer (0.765) Fetal (0.118) Inflammation (0.118)	pSPORT1
177	Gastr stinal (1.000)		Cancer (0.667) Inflammation (0.333)	pINCY
178	Repro ive (0.38) Nervou Gastr stinal (0.154)	31)	Cancer (0.385) Inflammation (0.385)	pINCY
179	Repr ive (0.500) Cardiov Gastr stinal (0.167)	lar (0.167)	Cancer (0.667) Fetal (0.167) Inflammation (0.167)	pBLUESCRIPT
180	Cardi ular (0.2 1) Repr Gastr stinal (0.154)	e (0.231)	Cancer (0.615) Inflammation (0.308) Fetal (0.154)	pINCY
181	Repro ctive (0.32) Gastroi Cardi ascular (0.130)	al (0.176)	Cancer (0.519) Inflammation (0.222) Fetal (0.157)	pINCY
182	Reproduc ive (0.320) Nervou Gastroin stinal (0.120)	180)	Cancer (0.580) Inflammation (0.160) Fetal (0.100)	pINCY
183	Gastr stinal (0. 67) Repr	ve (0.333)	Cancer (1.000)	pINCY
184	Urolo ic (0.667) D rmatolog	33)	Cancer (0.667) Fetal (0.333)	pSPORT1
185	Cardi ascular (0.5 0) Repr	ve (0.500)	Cancer (1.000)	pINCY
186	Repr ive (0.39) Develo Urolo 0.107)	al (0.107)	Cancer (0.607) Fetal (0.179) Inflammation (0.107)	pINCY
187	Cardi cular (0.4 0) Repr Gastr stinal (0.133)	ve (0.333)	Inflammation (0.467) Cancer (0.267) Fetal (0.267)	pSPORT1
188	Nerv 0.318) R oproducti	7) Urologic (0.136)	Cancer (0.636) Inflammation (0.136) Trauma (0.091)	pINCY

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissue ' ression (Fraction o	l)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
189	Cardio ular (0.500) Repro	(0.500)	Cancer (1.000)	pINCY
190	Reprod ve (0.318) Nervou Hemat etic/Immune (0.13	7)	Cancer (0.500) Fetal (0.227) Inflammation (0.227)	pINCY
191	Reprod ve (0.253) Cardiov Gastro tinal (0.147)	r (0.158)	Cancer (0.463) Inflammation (0.232) Fetal (0.200)	pINCY
192	Reprod ve (0.333) Gastroin Cardio ular (0.095)	al (0.286)	Cancer (0.571) Inflammation (0.333) Fetal (0.095)	pINCY
193	Reprod ve (0.304) Cardiov Gastroi tinal (0.130)	r (0.217)	Cancer (0.435) Inflammation (0.391) Fetal (0.174)	pINCY
194	Repro ve (0.312) Nervou Cardio ular (0.125)	8)	Cancer (0.438) Inflammation (0.250) Fetal (0.188)	pINCY
195	Devel ntal (1.000)		Fetal (1.000)	pINCY
196	Reprod ve (0.233) Cardiov Nervous (0.140)	r (0.209)	Cancer (0.605) Fetal (0.186) Inflammation (0.116)	pINCY
197	Reproductive (0.182) Gastroin Hematopoietic/Immune (0.13	al (0.136)	Cancer (0.477) Inflammation (0.341) Fetal (0.182)	pINCY
198	Gastro tinal (0.205) Repro Cardio ular (0.114)	ive (0.205)	Inflammation (0.341) Cancer (0.250) Fetal (0.227)	pINCY
199	Cardio ular (0.520) Repro Devel ntal (0.160)	(0.280)	Cancer (0.720) Fetal (0.200) Inflammation (0.080)	pINCY
200	Lung (0 8) Developmental (Muscu eletal (0.042)	)	Cancer (0.583) Fetal or Proliferating (0.292) Inflammation (0.167)	pBLUESCRIPT
201	Reprod e (0.57) Muscul Nervo .143) Ur logic (0.	al (0.143)	Cancer (0.429) Inflammation (0.571)	pSPORT1

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissue ression (Fraction o	l)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
202	Endoc 0.250) Nervous (0. Develo tal (0.125) Gastr Reprodu e (0.125)	Cardiovascular (0 125) inal (0.125)	Cancer (0 375) Inflammation (0.625) Fetal or Proliferating (0.125)	pSPORT1
203	Lung ( )		Fetal or Proliferating (1.000)	pINCY
204	Lung 0) Penis (0.500)		Cancer (0.500)	pINCY
205	Cardio lar (0.231) Derm ic (0.231) Repro ve (0.231)		Fetal or Proliferating (0.385) Cancer (0.308)	pINCY
206	Nervo 96) R roductiv 1) Gastroi inal (0.077)		Cancer (0.442) Neurological (0.192) Inflammation (0.231)	pINCY
207	Gastr inal (1.000)		Inflammation (1.000)	pINCY
208	Repro ve (0.300) Hemat c/Immune (0.200) Nervo 150)		Cancer (0 450) Inflammation (0.400) Fetal or Proliferating (0.250)	pSPORT1
209	Heart (0) Brain (0.500)		Neurological (0.500) Inflammation (0.500)	pINCY
210	Nervo 625) Re roductiv 0) Muscu eletal (0. 25)		Cancer (0.750) Fetal or Proliferating (0.250) Neurological (0.125)	pINCY
211	Nerv 261) Re roductiv 4) Gastr tinal (0.174)		Cancer (0.522) Fetal or Proliferating (0.174) Inflammation (0.130)	pSPORT1
212	Testi 0)		Inflammation (1.000)	pBLUESCRIPT
213	Nerv 400) Reproductiv 0) Gastr tinal (0.200)		Cancer (0.400) Inflammation (0.400) Neurological (0.200)	pBLUESCRIPT
214	Repro ve (0.476) Gastroi l (0.286) Cardi ular (0.0 5)		Cancer (0.714) Inflammation (0.286) Neurological (0.048)	pSPORT1

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissue Expression (Fraction	tal)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
215	Repro tive (0.28 4) Gastro Nervo 0.176) Hematopoi Cardi cular (0.108)	nal (0.216) mmune (0.108)	Cancer (0.486) Inflammation (0.351) Fetal or Proliferating (0.122)	pSPORT1
216	Uteru 500) Prostate (0.5(		Cancer (0.500) Inflammation (0.500)	pINCY
217	Nervo 0.429) Cardiovasc Gastr stinal (0.143) Her Repr tive (0.143)	(0.143) ietic/Immune (0.143)	Cancer (0.571) Inflammation (0.429) Fetal or Proliferating (0.285)	pSPORT1
218	Repr tive (0.450) Hema Nerv 0.100) Gastrointes	tic/Immune (0.200) (0.100)	Cancer (0.650) Inflammation (0.200) Fetal or Proliferating (0.050)	pINCY
219	Rep uctive (0.364) Cardi Nerv s (0.182)	lar (0.182)	Cancer (0.636) Fetal or Proliferating (0.182) Inflammation (0.273)	pINCY
220	Prost te (1.000)		Inflammation (1.000)	pSPORT1
221	Dev mental (0.333) Nerv Repr tive (0.333)	.333)	Cancer (0.667) Fetal or Proliferating (0.667)	pSPORT1
222	Repr tive (0.3 3) Hema Nerv (0.098) Cardiovasc	tic/Immune (0.180) 0.098)	Cancer (0.508) Inflammation (0.344) Fetal or Proliferating (0.066)	pSPORT1
223	End e (0.333) Gastroint Repr tive (0.333)	l (0.333)	Cancer (1.000)	pINCY
224	Car scular (0.200) Dev Gastrointestinal (0.200) Rep Urologic (0.200)	ental (0.200) ive (0.200)	Cancer (0.800) Fetal or Proliferating (0.200)	pINCY
225	Lung (1 00)		Cancer (1.000)	pINCY
226	Rep tive (0.302) Hema Car cular (0.111)	tic/Immune (0.254)	Cancer (0.381) Inflammation (0.381) Fetal or Proliferating (0.286)	pSPORT1

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissue [illegible]pression (Fraction [illegible]tal)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
227	Lymph[illegible]tes (1.000)	Inflammation (1.000)	pINCY
228	Cardi[illegible]cular (0.531) Repr[illegible]ve (0.250) Urolo[illegible]0.094)	Cancer (0.656) Inflammation (0.250) Fetal or Proliferating (0.094)	pINCY
229	Repro[illegible]ive (0.333) Cardio[illegible]r (0.167) Gastr[illegible]stinal (0.167) Endo[illegible](0.167) Hema[illegible]ietic/Immune (0.1[illegible]	Cancer (0.500) Fetal or Proliferating (0.167) Inflammation (0.333)	pINCY
230	Hema[illegible]ietic/Immune (0.5[illegible] productive (0.500)	Cell Proliferation (0.500) Inflammation (0.500)	pBLUESCRIPT
231	Cardi[illegible]cular (0.333) Nerv[illegible].333) Devel[illegible]ental (0.167)	Cancer (0.500) Cell Proliferation (0.333) Inflammation (0.167)	pINCY
232	Gastrointestinal (0.[illegible]38) Repr[illegible]ive (0.062)	Cancer (0.500) Inflammation (0.500)	pINCY
233	Nervous (0.324) Reproductiv[illegible]235) Hematopoietic/Imm[illegible]une (0.11	Cancer (0.456) Inflammation (0.235) Trauma (0.147)	pINCY
234	Nervou[illegible].255) Reproductiv[illegible]255) Musc[illegible]keletal (0.182)	Cancer (0.545) Inflammation (0.255) Trauma (0.109)	pINCY
235	Musculoskeletal (0.308) Rep[illegible]ive (0.231) Gastrointestinal (0.[illegible]54)	Cancer (0.538) Inflammation (0.231) Trauma (0.154)	pINCY
236	Nervous (1.000)	Cancer (1.000)	pINCY
237	Gastrointestinal (0.429) Hematopoietic/Imm[illegible]une (0.1[illegible]rvous (0.143)	Cancer (0.571) Cell Proliferation (0.143) Trauma (0.143)	pINCY
238	Reproductive (0.25[illegible]) Gastro[illegible]al (0.160) Nerv[illegible](0.128)	Cancer (0.453) Inflammation (0.241) Cell Proliferation (0.175)	pINCY
239	Nervo[illegible].333) D[illegible]rmatolog[illegible].7) Endo[illegible](0.167)	Trauma (0.333) Cancer (0.167) Cell Proliferation (0.167)	pINCY

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissue Expression (Fraction ...tal)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
240	Nerv 0.273) Reproducti ...27) End (0.136)	Cancer (0.545) Cell Proliferation (0.182) Inflammation (0.182)	pINCY
241	Repr tive (0.273) Hema ietic/Immune (0.1 ...rologic (0.182)	Cancer (0.455) Cell Proliferation (0.273) Inflammation (0.273)	pINCY
242	Endo (1.000)	Trauma (1.000)	pSPORT1
243	Repr tive (1.000)	Cancer (1.000)	pINCY
244	Hem ietic/Immune (0.5 Musculo keletal (0.182) Car ...cular (0.091)	Inflammation (0.636) Trauma (0.182) Cancer (0.091)	pINCY
245	Hematopoietic/Immune (0.4 Musculoskeletal (0.300) Car ...cular (0.150)	Inflammation (0.650) Cancer (0.300)	pINCY
246	Urol (1.000)	Cancer (0.500) Cell Proliferation (0.500)	pINCY
247	Nerv 0.292) Reproducti ...22) Mus keletal (0.125)	Cell Proliferation (0.625) Inflammation/Trauma (0.181)	pSPORT1
248	Repro ve (0.211) Develo ...al (0.132) Nerv 0.132)	Cell Proliferation (0.658) Inflammation/Trauma (0.184)	pSPORT1
249	Nervous (0.500) Gastrointest ...0.300) Hema ietic/Immune (0.1	Cell Proliferation (0.900) Inflammation/Trauma (0.300)	pSPORT1
250	Cardiovascular (0.209) Gastr ...stinal (0.140) Hematopoietic/Im une (0.1	Cell Proliferation (0.605) Inflammation/Trauma (0.256)	pINCY
251	Nervous (0.308) Cardiovascu ...0.154) Gastrointestinal (0.154)	Cell Proliferation (0.616) Inflammation/Trauma (0.269)	pINCY
252	Repr ive (1.000)	Cell Proliferation (1.000)	pSPORT1

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissue [illegible]pression (Fraction c[illegible]al)	Disease/Condition-Specific Expression (Total of Fraction)	Vector
253	Repro[illegible]ive (0.32[illegible]) Nervou[illegible].162) Gastr[illegible]stinal (0.113)	Cell Proliferation (0.641) Inflammation/Trauma (0.197)	pSPORT1
254	Repro[illegible]ive (0.315) Nervou[illegible]96) Deve[illegible]ntal (0.093)	Cell Proliferation (0.630) Inflammation/Trauma (0.278)	pSPORT1
255	Nervo[illegible]0.211) Reproductiv[illegible].211) Gastr[illegible]tinal (0.158)	Cell Proliferation (0.579) Inflammation/Trauma (0.298)	pINCY
256	Repro[illegible]ive (0.25[illegible]) Gastro[illegible]nal (0.148) Hema[illegible]ietic/Immune (0.1[illegible]	Cell Proliferation (0.705) Inflammation/Trauma (0.193)	pINCY
257	Hema[illegible]ietic/Immune (1.0[illegible]	Cell Proliferation (0.400) Inflammation/Trauma (0.600)	pINCY
258	Cardi[illegible]cular (0.333) Repro[illegible]tive (0.333) Develo[illegible]ental (0.167)	Cell Proliferation (0.833) Inflammation/Trauma (0.333)	pBLUESCRIPT
259	Cardi[illegible]cular (0.333) Repro[illegible]ve (0.250) Deve[illegible]ntal (0.167)	Cell Proliferation (0.625) Inflammation/Trauma (0.208)	pINCY
260	Endocrine (0.500) Cardiovas[illegible] (0.250) Nervous (0.250)	Cell Proliferation (0.750) Inflammation/Trauma (0.500)	pINCY
261	Reproductive (0.252) Cardio[illegible]ar (0.155) Hematopo[illegible]etic/Immune (0.1[illegible]	Cell Proliferation (0.728) Inflammation/Trauma (0.194)	pINCY
262	Reproduc[illegible]ive (0.2[illegible]) Cardio[illegible]ar (0.177) Nervous (0.145)	Cell Proliferation (0.742) Inflammation/Trauma (0.210)	pINCY
263	Reproductive (0.267) Cardio[illegible]ar (0.160) Hematopoietic/Immune (0.1[illegible]	Cell Proliferation (0.654) Inflammation/Trauma (0.193)	pINCY
264	Nervous (0.229) H[illegible]matopoi[illegible]une (0.200) Repro[illegible]tive (0.200)	Cell Proliferation (0.743) Inflammation/Trauma (0.286)	pINCY
265	Hematop[illegible]ietic/Immune (0.3[illegible] [illegible]ointestinal (0.167) Nervous (0.133)	Cell Proliferation (0.600) Inflammation/Trauma (0.333)	pINCY

TABLE 3 (cont.)

Nucleotide SEQ ID NO:	Tissu	pression (Fraction	otal)	Disease/Condition-Specific Expression (Total of Fraction)		Vector
266	Nerv Cardi	0.290) R cular (0.129)	productiv 258)	Cell Proliferation (0.677) Inflammation/Trauma (0.194)		pINCY
267	Repr Cardi	tive (0.261) Hemat cular (0.087)	tic/Immune (0.217)	Cell Proliferation (0.652) Inflammation/Trauma (0.391)		pINCY
268	Gastr Hem	stinal (0.227) Rep ietic/Immune (0.1	ive (0.193)	Cell Proliferation (0.731) Inflammation/Trauma (0.227)		pSPORT1

TABLE 4

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
135	443531	GNOT03	The library was constructed using RNA isolated from plastic adherent mononuclear cells isolated from buffy coat units obtained from unrelated male and female donors.
136	632860	TGMT01	The library was constructed using RNA isolated from peripheral blood granulocytes collected by density gradient centrifugation through Ficoll-Hypaque. The cells were isolated from buffy coat units obtained from 20 unrelated male and female donors. Cells were cultured in 10 nM GM-CSF for 1 hour before washing and harvesting for RNA preparation.
137	670010	LNOT01	The library was constructed using RNA isolated from the cerebellum tissue of a 69-year-old Caucasian male who died from chronic obstructive pulmonary disease. Patient history included myocardial infarction, hypertension, and osteoarthritis.
138	726498	OAT01	The library was constructed using RNA isolated from the knee synovial membrane tissue of an 82-year-old female with osteoarthritis.
139	795064	RNOT03	The library was constructed using RNA isolated from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, cerebrovascular disease, breast cancer, and uterine cancer.
140	924925	NOT04	The library was constructed using RNA isolated from the brain tissue of a 44-year-old Caucasian male with a cerebral hemorrhage. The tissue, which contained coagulated blood, came from the choroid plexus of the right anterior temporal lobe. Family history included coronary artery disease and myocardial infarction.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	lone ID	ibrary	Library Description
141	962390	TTUT03	The library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.
142	1259405	ITUT03	The library was constructed using RNA isolated from brain meningioma tissue removed from a 35-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a benign neoplasm in the right cerebellopontine angle of the brain. Patient history included hypothyroidism. Family history included myocardial infarction and breast cancer.
143	1297384	NOT07	The library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, atherosclerotic coronary artery disease, and type II diabetes.
144	1299627	NOT07	The library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, atherosclerotic coronary artery disease, and type II diabetes.
145	1306026	CNOT02	The library was constructed using RNA isolated from the placental tissue of a Hispanic female fetus, who was prematurely delivered at 21 weeks' gestation. Serologies of the mother's blood were positive for CMV (cytomegalovirus).

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
146	1316219	DTUT02	The library was constructed using RNA isolated from bladder tumor tissue removed from an 80-year-old Caucasian female during a radical cystectomy and lymph node excision. Pathology indicated grade 3 invasive transitional cell carcinoma. Family history included osteoarthritis and atherosclerosis.
147	1329031	CNOT07	The library was constructed using RNA isolated from the pancreatic tissue of a Caucasian male fetus, who died at 23 weeks' gestation.
148	83050	NOT02	The library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
149	1514160	CTUT01	The library was constructed using RNA isolated from pancreatic tumor tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Family history included cardiovascular disease, type II diabetes, and stomach cancer.
150	1603403	GNOT15	The library was constructed using RNA isolated from lung tissue removed from a 69-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated residual grade 3 invasive squamous cell carcinoma. Patient history included acute myocardial infarction, prostatic hyperplasia, and malignant skin neoplasm. Family history included cerebrovascular disease, type I diabetes, acute myocardial infarction, and arteriosclerotic coronary disease.
151	1652303	STUT08	The library was constructed using RNA isolated from prostate tumor tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated an adenocarcinoma (Gleason grade 3+4). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst. Family history included tuberculosis, cerebrovascular disease, and arteriosclerotic coronary artery disease.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
152	1693358	...NOT23	The library was constructed using RNA isolated from diseased colon tissue removed from a 16-year-old Caucasian male during a total colectomy with abdominal/perineal resection. Pathology indicated gastritis and pancolonitis consistent with the acute phase of ulcerative colitis.There was only mild involvement of the ascending and sigmoid colon, and no significant involvement of the cecum, rectum, or terminal ileum. Family history included irritable bowel syndrome.
153	1707711	...DNOT02	The library was constructed using RNA isolated from duodenal tissue of a 8-year-old Caucasian female, who died from head trauma. Serology was positive for cytomegalovirus (CMV).
154	1738735	...NOT22	The library was constructed using RNA isolated from colon tissue removed from a 56-year-old Caucasian female with Crohn's disease during a partial resection of the small intestine. Pathology indicated Crohn's disease of the ileum and ileal-colonic anastomosis, causing a fistula at the anastomotic site that extended into pericolonic fat. The ileal mucosa showed linear and puncture ulcers with intervening normal tissue. Previous surgeries included a partial ileal resection and permanent ileostomy. Family history included irritable bowel syndrome.
155	1749147	...TUT02	The library was constructed using RNA isolated from stomach tumor tissue obtained from a 68-year-old Caucasian female during a partial gastrectomy. Pathology indicated a malignant lymphoma of diffuse large-cell type. Patient history included thalassemia. Family history included acute leukemia, malignant neoplasm of the esophagus, malignant stomach neoplasm, and atherosclerotic coronary artery disease.
156	1817722	...NOT20	The library was constructed using RNA isolated from diseased prostate tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma.
157	1831290	...AZT01	The library was constructed using 1 microgram of polyA RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
158	831477	TH' T01	The library was constructed using 1 microgram of polyA RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia.
159	1841607	NNOT07	The library was constructed using RNA isolated from colon tissue removed from a 60-year-old Caucasian male during a left hemicolectomy.
160	852391	GFET03	The library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
161	854555	AZT01	Library was constructed using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated for three days with 0.35 micromolar 5-aza-2'-deoxycytidine (AZT).
162	855755	SNOT18	The library was constructed using RNA isolated from diseased prostate tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated adenofibromatous hyperplasia. This tissue was associated with a grade 3 transitional cell carcinoma. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
163	1861434	NOT19	The library was constructed using RNA isolated from diseased prostate tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+3). The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli and thrombophlebitis. Family history included benign hypertension, multiple myeloma, hyperlipidemia and rheumatoid arthritis.
164	872334	OT02	The library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).
165	1877230	OT03	The library was constructed using RNA isolated from white blood cells of a 27-year-old female with blood type A+. The donor tested negative for cytomegalovirus (CMV).

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID		rary	Library Description
166	877885	L	NOT03	The library was constructed using RNA isolated from white blood cells of a 27-year-old female with blood type A+. The donor tested negative for cytomegalovirus (CMV).
167	1889269	P	TUT07	The library was constructed using RNA isolated from bladder tumor tissue removed from the anterior bladder wall of a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated a grade 3 transitional cell carcinoma in the left lateral bladder. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
168	1890243	P	TUT07	The library was constructed using RNA isolated from bladder tumor tissue removed from the anterior bladder wall of a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated a grade 3 transitional cell carcinoma in the left lateral bladder. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
169	1900433	P	TUT06	The library was constructed using RNA isolated from bladder tumor tissue removed from the posterior bladder wall of a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated grade 3 transitional cell carcinoma in the left lateral bladder wall. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
170	1909441	C	TUT01	The library was constructed using RNA isolated from a soft tissue tumor removed from the clival area of the skull of a 30-year-old Caucasian female. Pathology indicated chondroid chordoma with neoplastic cells reactive for keratin.
171	1932226	C	NOT16	The library was constructed using RNA isolated from sigmoid colon tissue removed from a 62-year-old Caucasian male during a sigmoidectomy and permanent colostomy.
172	1932647	C	NOT16	The library was constructed using RNA isolated from sigmoid colon tissue removed from a 62-year-old Caucasian male during a sigmoidectomy and permanent colostomy.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO	ne ID	ary	Library Description
173	2124245	B OT07	The library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, atherosclerotic coronary artery disease, and type II diabetes.
174	132626	O NOT03	The library was constructed using RNA isolated from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, cerebrovascular disease, breast cancer, and uterine cancer.
175	280639	P NON01	The library was constructed and normalized from 4.4 million independent clones from the PROSNOT11 library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.
176	292356	P NON01	The library was constructed and normalized from 4.88 million independent clones from the BRAINOT03 library. Starting RNA was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
177	349310	C JCT01	The library was constructed using RNA isolated from diseased sigmoid colon tissue obtained from a 70-year-old Caucasian male during colectomy with permanent ileostomy. Pathology indicated chronic ulcerative colitis. Patient history included benign neoplasm of the colon. Family history included atherosclerotic coronary artery disease and myocardial infarctions.
178	373227	A OT07	The library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
179	457682	ANOT01	The library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.
180	2480426	NOT01	The library was constructed using RNA isolated from an aortic smooth muscle cell line derived from the explanted heart of a male during a heart transplant.
181	2503743	TUT01	The library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed mullerian tumor present in the sigmoid mesentery at two sites.
182	2537684	TUT01	The library was constructed using RNA isolated from rib tumor tissue removed from a 16-year-old Caucasian male during a rib osteotomy and a wedge resection of the lung. Pathology indicated a metastatic grade 3 (of 4) osteosarcoma, forming a mass involving the chest wall.
183	2593853	TUT02	The library was constructed using RNA isolated from ovarian tumor tissue removed from a 51-year-old Caucasian female during an exploratory laparotomy, total abdominal hysterectomy, salpingo-oophorectomy, and an incidental appendectomy. Pathology indicated mucinous cystadenoma presenting as a multiloculated neoplasm involving the entire left ovary. The right ovary contained a follicular cyst and a hemorrhagic corpus luteum. The uterus showed proliferative endometrium and a single intramural leiomyoma. The peritoneal biopsy indicated benign glandular inclusions consistent with endosalpingiosis. Family history included atherosclerotic coronary artery disease, benign hypertension, breast cancer, and uterine cancer.
184	2622354	NOT02	The library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is a human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	ne ID	rary	Library Description
185	41377	TUT08	The library was constructed using RNA isolated from lung tumor tissue removed from a 63-year-old Caucasian male during a right upper lobectomy with fiberoptic bronchoscopy. Pathology indicated a grade 3 adenocarcinoma. Patient history included atherosclerotic coronary artery disease, an acute myocardial infarction, rectal cancer, an asymptomatic abdominal aortic aneurysm, and cardiac dysrhythmia. Family history included congestive heart failure, stomach cancer, and lung cancer, type II diabetes, atherosclerotic coronary artery disease, and an acute myocardial infarction.
186	4857	NOT19	The library was constructed using RNA isolated from kidney tissue removed a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. Pathology for the associated tumor tissue indicated a grade 1 renal cell carcinoma within the upper pole of the left kidney. Patient history included malignant melanoma of the abdominal skin, benign neoplasm of colon, cerebrovascular disease, and umbilical hernia. Family history included myocardial infarction, atherosclerotic coronary artery disease, cerebrovascular disease, prostate cancer, myocardial infarction, and atherosclerotic coronary artery disease.
187	2758485	ZS08	The subtracted THP-1 promonocyte cell line library was constructed using 5.76 million clones from a 5-aza-2'-deoxycytidine (AZT) treated THP-1 cell library. Starting RNA was made from THP-1 promonocyte cells treated for three days with 0.8 micromolar AZT. The library was oligo(dT)-primed, and cDNAs were cloned directionally into the pSPORT1 vectoring system using SalI (5') and NotI (3'). The hybridization probe for subtraction was derived from a similarly constructed library, made from 1 microgram of polyA RNA isolated from untreated THP-1 cells. 5.76 million clones from the AZ-treated THP-1 cell library were then subjected to two rounds of subtractive hybridization with 5 million clones from the untreated THP-1 cell library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (Nucl. Acids Res. (1991) 19:1954) and Bonaldo et al. (Genome Res (1996) 6: 791-806).
188	63296	OT12	The library was constructed using RNA isolated from diseased breast tissue removed from a 32-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated nonproliferative fibrocystic disease. Family history included benign hypertension and atherosclerotic coronary artery disease.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
189	779436	OVUT03	The library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 52-year-old mixed ethnicity female during a total abdominal hysterectomy, bilateral salpingo-oophorectomy, peritoneal and lymphatic structure biopsy, regional lymph node excision, and peritoneal tissue destruction. Pathology indicated an invasive grade 3 (of 4) seroanaplastic carcinoma forming a mass in the left ovary. The endometrium was atrophic. Multiple (2) leiomyomata were identified, one subserosal and 1 intramural. Pathology also indicated a metastatic grade 3 seroanaplastic carcinoma involving the omentum, cul-de-sac peritoneum, left broad ligament peritoneum, and mesentery colon. Patient history included breast cancer, chronic peptic ulcer, and joint pain. Family history included colon cancer, cerebrovascular disease, breast cancer, type II diabetes, esophagus cancer, and depressive disorder.
190	708528	BLADUT08	The library was constructed using RNA isolated from bladder tumor tissue removed from a 72-year-old Caucasian male during a radical cystectomy and prostatectomy. Pathology indicated an invasive grade 3 (of 3) transitional cell carcinoma in the right bladder base. Family history included myocardial infarction, cerebrovascular disease, brain cancer, and myocardial infarction.
191	809230	BLADUT08	The library was constructed using RNA isolated from bladder tumor tissue removed from a 72-year-old Caucasian male during a radical cystectomy and prostatectomy. Pathology indicated an invasive grade 3 (of 3) transitional cell carcinoma in the right bladder base. Patient history included pure hypercholesterolemia and tobacco abuse. Family history included myocardial infarction, cerebrovascular disease, brain cancer, and myocardial infarction.
192	816821	BRSTNOT14	The library was constructed using RNA isolated from breast tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. The tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included a benign colon neoplasm, hyperlipidemia, and cardiac dysrhythmia. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
193	72(	TNOT14	The library was constructed using RNA isolated from breast tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. The tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included a benign colon neoplasm, hyperlipidemia, and cardiac dysrhythmia. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease.
194	923163	TNOT04	The library was constructed using RNA isolated from diseased ileum tissue obtained from a 26-year-old Caucasian male during a partial colectomy, permanent colostomy, and an incidental appendectomy. Pathology indicated moderately to severely active Crohn's disease. Family history included enteritis of the small intestine.
195	49822	NFET01	The library was constructed using RNA isolated from kidney tissue removed from a Caucasian female fetus, who died at 17 weeks' gestation from anencephalus.
196	92192	NFET02	The library was constructed using RNA isolated from kidney tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart and died at 23 weeks' gestation.
197	092453	NFET02	The library was constructed using RNA isolated from kidney tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart and died at 23 weeks' gestation.
198	044710	ANOT01	The library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending coronary artery, atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, and left ventricular dysfunction. Previous surgeries included cardiac catheterization. Family history included atherosclerotic coronary artery disease.

TABLE 4 (cont.)

Polynucleot SEQ ID NO:	one ID	rary	Library Description
199	120415	L TUT13	The library was constructed using RNA isolated from tumorous lung tissue removed from the right upper lobe of a 47-year-old Caucasian male during a segmental lung resection. Pathology indicated invasive grade 3 (of 4) adenocarcinoma. Family history included atherosclerotic coronary artery disease, and type II diabetes.
200	26758	L NOT01	The library was constructed at Stratagene using RNA isolated from the lung tissue of a 72-year-old male.
201	4760	C NOT01	The library was constructed using RNA isolated from the cerebellum tissue of a 69-year-old Caucasian male who died from chronic obstructive pulmonary disease. Patient history included myocardial infarction, hypertension, and osteoarthritis.
202	229438	I UT01	The library was constructed using RNA isolated from brain tumor tissue removed from a 50-year-old Caucasian female during a frontal lobectomy. Pathology indicated recurrent grade 3 oligoastrocytoma with focal necrosis and extensive calcification. Patient history included a speech disturbance and epilepsy. The patient's brain had also been irradiated with a total dose of 5,082 cyg (Fraction 8). Family history included a brain tumor.
203	36935	I GFET03	The library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus who died at 20 weeks' gestation.
204	[illegible]	L NOT12	The library was constructed using RNA isolated from lung tissue removed from a 78-year-old Caucasian male during a segmental lung resection and regional lymph node resection. Pathology indicated fibrosis pleura was puckered, but not invaded. Pathology for the associated tumor tissue indicated an invasive pulmonary grade 3 adenocarcinoma. Patient history included cerebrovascular disease, arteriosclerotic coronary artery disease, thrombophlebitis, chronic obstructive pulmonary disease, and asthma. Family history included intracranial hematoma, cerebrovascular disease, arteriosclerotic coronary artery disease, and type I diabetes.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
205	[illegible]	TUT01	The library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.
206	910668	TUT01	The library was constructed using RNA isolated from a soft tissue tumor removed from the clival area of the skull of a 30-year-old Caucasian female. Pathology indicated chondroid chordoma with neoplastic cells reactive for keratin.
207	55143	NNOT01	The library was constructed using RNA isolated from mesentery fat tissue obtained from a 71-year-old Caucasian male during a partial colectomy and permanent colostomy. Family history included atherosclerotic coronary artery disease, myocardial infarction, and extrinsic asthma.
208	961637	NOT04	The library was constructed using RNA isolated from breast tissue removed from a 62-year-old East Indian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 ductal carcinoma. Patient history included benign hypertension, hyperlipidemia, and hematuria. Family history included cerebrovascular and cardiovascular disease, hyperlipidemia, and liver cancer.
209	1990762	PNOT02	The library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
210	1994131	PNOT02	The library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.

TABLE 4 (cont.)

Polynucle SEQ ID NO:	ne ID	rary	Library Description
211	[illegible]	TTUT03	The library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.
212	09035	NOT03	The library was constructed using polyA RNA isolated from testicular tissue removed from a 37-year-old Caucasian male who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
213	009152	NOT03	The library was constructed using polyA RNA isolated from testicular tissue removed from a 37-year-old Caucasian male who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
214	061752	RNOT03	The library was constructed using RNA isolated from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.
215	193	RNOT03	The library was constructed using RNA isolated from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
216		NOT08	The library was constructed using RNA isolated from uterine tissue removed from a 35-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated that the endometrium was secretory phase with a benign endometrial polyp 1 cm in diameter. The cervix showed mild chronic cervicitis. Family history included atherosclerotic coronary artery disease and type II diabetes.
217	101278	ITUT02	The library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
218	121353	NOT07	The library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, cardiovascular disease, and type II diabetes.
219	241736	CTUT02	The library was constructed using RNA isolated from pancreatic tumor tissue removed from a 45-year-old Caucasian female during radical pancreaticoduodenectomy. Pathology indicated a grade 4 anaplastic carcinoma. Family history included benign hypertension, hyperlipidemia and atherosclerotic coronary artery disease.
220	271935	SNON01	This normalized prostate library was constructed from 4.4 M independent clones from the PROSNOT11 library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl Acad Sci USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
221	0534	NOT05	The library was constructed using RNA isolated from breast tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated multicentric invasive grade 4 lobular carcinoma. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular and cardiovascular disease, breast and prostate cancer, and type I diabetes.
222	0390	NOT05	The library was constructed using RNA isolated from breast tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated multicentric invasive grade 4 lobular carcinoma. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular and cardiovascular disease, breast and prostate cancer, and type I diabetes.
223	2097805	TUT05	The library was constructed RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma.
224	2646362	GTUT11	The library was constructed using RNA isolated from lung tumor tissue removed from the right lower lobe a 57-year-old Caucasian male during a segmental lung resection. Pathology indicated an infiltrating grade 4 squamous cell carcinoma. Multiple intrapulmonary peribronchial lymph nodes showed metastatic squamous cell carcinoma. Patient history included a benign brain neoplasm and tobacco abuse. Family history included spinal cord cancer, type II diabetes, cerebrovascular disease, and malignant prostate neoplasm.
225	5718	GTUT09	The library was constructed using RNA isolated from lung tumor tissue removed from a 68-year-old Caucasian male during segmental lung resection. Pathology indicated invasive grade 3 squamous cell carcinoma and a metastatic tumor. Patient history included type II diabetes, thyroid disorder, depressive disorder, hyperlipidemia, esophageal ulcer, and tobacco use.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone	Library	Library Description
226		AZS08	This subtracted THP-1 promonocyte cell line library was constructed using 5.76 million clones from a 5-aza-2'-deoxycytidine (AZ) treated THP-1 cell library. Starting RNA was made from THP-1 promonocyte cells treated for three days with 0.8 micromolar AZ. The hybridization probe for subtraction was derived from a similarly constructed library, made from RNA isolated from untreated THP-1 cells. 5.76 million clones from the AZ-treated THP-1 cell library were then subjected to two rounds of subtractive hybridization with 5 million clones from the untreated THP-1 cell library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954, and Bonaldo et al., Genome Research (1996) 6:791. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia.
227		NOT03	The library was constructed using RNA isolated from nonactivated Th1 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-12 and B7-transfected COS cells.
228	116250	GTUT13	The library was constructed using RNA isolated from tumorous lung tissue removed from the right upper lobe of a 47-year-old Caucasian male during a segmental lung resection. Pathology indicated invasive grade 3 (of 4) adenocarcinoma. Family history included atherosclerotic coronary artery disease, and type II diabetes.
229	12963	GTUT12	The library was constructed using RNA isolated from tumorous lung tissue removed from a 70-year-old Caucasian female during a lung lobectomy of the left upper lobe. Pathology indicated grade 3 (of 4) adenocarcinoma and vascular invasion. Patient history included tobacco abuse, depressive disorder, anxiety state, and skin cancer. Family history included cerebrovascular disease, congestive heart failure, colon cancer, depressive disorder, and primary liver.
230	07632	C1NOT01	The library was constructed using RNA isolated from the HMC-1 human mast cell line derived from a 52-year-old female. Patient history included mast cell leukemia.
231		GFET03	The library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus who died at 20 weeks' gestation.
232		NOT13	The library was constructed using RNA isolated from ascending colon tissue of a 28-year-old Caucasian male with moderate chronic ulcerative colitis.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	[illegible]	Library	Library Description
233	396972	AITUT08	The library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 47-year-old Caucasian male during excision of cerebral meningeal tissue. Pathology indicated grade 4 fibrillary astrocytoma with focal tumoral radionecrosis. Patient history included cerebrovascular disease, deficiency anemia, hyperlipidemia, epilepsy, and tobacco use. Family history included cerebrovascular disease and malignant prostate neoplasm.
234	[illegible]	IBST01	The library was constructed using RNA isolated from ileum tissue removed from an 18-year-old Caucasian female during bowel anastomosis. Pathology indicated Crohn's disease of the ileum. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.
235	[illegible]	DNOT03	The library was constructed using RNA isolated from lymph node tissue removed from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. This tissue was extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, and congestive heart failure.
236	1547884	STUT09	The library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. Patient history included lung neoplasm, and benign hypertension. Family history included malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer.
237	1661144	TNOT09	The library was constructed using RNA isolated from breast tissue removed from a 45-year-old Caucasian female during unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated invasive nuclear grade 2-3 adenocarcinoma. Patient history included valvuloplasty of mitral valve and rheumatic heart disease. Family history included cardiovascular disease and type II diabetes.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:		Library	Library Description
238	1665	SNOT15	The library was constructed using RNA isolated from diseased prostate tissue removed from a 66-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma (Gleason grade 2+3). The patient presented with elevated prostate specific antigen (PSA). Family history included prostate cancer, secondary bone cancer, and benign hypertension.
239		TTUT08	The library was constructed using RNA isolated from breast tumor tissue removed from a 45-year-old Caucasian female during unilateral extended simple mastectomy. Pathology indicated invasive nuclear grade 2-3 adenocarcinoma. Patient history included valvuloplasty of mitral valve and rheumatic heart disease. Family history included cardiovascular disease and type II diabetes.
240	650265	NOT14	The library was constructed using RNA isolated from breast tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma. Patient history included a benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included cardiovascular and cerebrovascular disease and colon, ovary and lung cancer.
241		NNOT19	The library was constructed using RNA isolated from kidney tissue removed a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. Pathology for the associated tumor tissue indicated grade 1 renal cell carcinoma within the upper pole of the left kidney. Patient history included malignant melanoma of the abdominal skin, benign neoplasm of colon, cerebrovascular disease, and umbilical hernia. Family history included myocardial infarction, atherosclerotic coronary artery disease, cerebrovascular disease, and prostate cancer.
242		ENON04	The normalized adrenal gland library was constructed from 1.36 x 1e6 independent clones from an adrenal tissue library. Starting RNA was made from adrenal gland tissue removed from a 20-year-old Caucasian male who died from head trauma. The library was normalized in two rounds using conditions adapted from Soares et al. (PNAS (1994) 91:9228-9232) and Bonaldo et al. (Genome Res (1996) 6: 791-806) using a significantly longer (48-hours/round) reannealing hybridization period.

TABLE 4 (cont.)

Polynucleotide SEQ ID NO:	Clone ID	Library	Library Description
243	[illegible]	[illegible]TNOT18	The library was constructed using RNA isolated from diseased breast tissue removed from a 57-year-old Caucasian female during a unilateral simple extended mastectomy. Pathology indicated mildly proliferative breast disease. Patient history included breast cancer and osteoarthritis. Family history included type II diabetes, gallbladder and breast cancer, and chronic lymphocytic leukemia.
244	[illegible]	[illegible]GNOT27	The library was constructed using RNA isolated from lung tissue removed from a 17-year-old Hispanic female.
245	[illegible]	[illegible]NDTNT01	The library was constructed using RNA isolated from treated dendritic cells from peripheral blood.
246	[illegible]	[illegible]NNOT26	The library was constructed using RNA isolated from left kidney medulla and cortex tissue removed from a 53-year-old Caucasian female during a nephroureterectomy. Pathology for the associated tumor tissue indicated grade 2 renal cell carcinoma involving the lower pole of the kidney. Patient history included hyperlipidemia, cardiac dysrhythmia, menorrhagia, cerebrovascular disease, atherosclerotic coronary artery disease, and tobacco abuse. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.
247	[illegible]	[illegible]TNOT03	The library was constructed using RNA isolated from diseased breast tissue removed from a 54-year-old Caucasian female during a bilateral radical mastectomy. Pathology for the associated tumor tissue indicated residual invasive grade 3 mammary ductal adenocarcinoma. Patient history included kidney infection and condyloma acuminatum. Family history included benign hypertension, hyperlipidemia and a malignant neoplasm of the colon.
248	[illegible]	[illegible]CNOT04	The library was constructed using RNA isolated from the pancreatic tissue of a 5-year-old Caucasian male who died in a motor vehicle accident.
249	[illegible]	[illegible]ITUT03	The library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe a 17-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a grade 4 fibrillary giant and small-cell astrocytoma. Family history included benign hypertension and cerebrovascular disease.
250	[illegible]	[illegible]CNOT03	The library was constructed using RNA isolated from dermal microvascular endothelial cells removed from a neonatal Caucasian male.

TABLE 4 (cont.)

Polynucle[illegible]e SEQ ID [illegible]:	[illegible]lone ID	Library	Library Description
251	[illegible]2212530	[illegible]NTFET03	The library was constructed using RNA isolated from small intestine tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
252	[illegible]253036	[illegible]ARTUT01	The library was constructed using RNA isolated from ovarian tumor tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology indicated grade 2 mucinous cystadenocarcinoma involving the entire left ovary. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.
253	2280161	[illegible]NON01	The normalized prostate library was constructed from 4.4 M independent clones from the PROSNOT11 library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.
254	[illegible]287485	[illegible]NON01	The library was constructed and normalized from 4.88 million independent clones from the BRAINOT03 library. RNA was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
255	[illegible]380344	[illegible]NOT01	The library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
256	2383171	[illegible]NOT01	The library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
257	2396046	[illegible]P1AZT01	The library was constructed using RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202)is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia.
258	[illegible]15658[illegible]	[illegible]NOT01	The library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.

TABLE 4 (cont.)

Polynucleotide SEQ ID	Clone ID	Library	Library Description
259	184813	[illegible]NRTUT01	The library was constructed using RNA isolated from rib tumor tissue removed from a 16-year-old Caucasian male during a rib osteotomy and a wedge resection of the lung. Pathology indicated a metastatic grade 3 (of 4) osteosarcoma, forming a mass involving the chest wall.
260	193851	[illegible]TUT05	The library was constructed RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma.
261	2495719	[illegible]ETUT05	The library was constructed RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma.
262	514153	[illegible]LANOT01	The library was constructed using RNA isolated from diseased gallbladder tissue removed from a 53-year-old Caucasian female during a cholecystectomy. Pathology indicated mild chronic cholecystitis and cholelithiasis with approximately 150 mixed gallstones. Family history included benign hypertension.
263	655183	[illegible]MNOT04	The library was constructed using RNA isolated from thymus tissue removed from a 3-year-old Caucasian male, who died from anoxia. Serologies were negative. The patient was not taking any medications.
264	2848362	[illegible]STTUT13	The library was constructed using RNA isolated from breast tumor tissue removed from the right breast of a 46-year-old Caucasian female during a unilateral extended simple mastectomy with breast reconstruction. Pathology indicated an invasive grade 3 adenocarcinoma, ductal type with apocrine features and greater than 50% intraductal component. Patient history included breast cancer.
265	3499[illegible]	[illegible]TTUT13	The library was constructed using RNA isolated from breast tumor tissue removed from the right breast of a 46-year-old Caucasian female during a unilateral extended simple mastectomy with breast reconstruction. Pathology indicated an invasive grade 3 adenocarcinoma, ductal type with apocrine features and greater than 50% intraductal component. Patient history included breast cancer.

TABLE 4 (cont.)

Polynucleotide SEQ ID	Clone ID	Library	Library Description
266	899137	CNOT01	The library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus, infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
267	3622	GDIT01	The library was constructed using RNA isolated from diseased cartilage tissue. Patient history included osteoarthritis.
268	2095	NNON03	The normalized colon library was constructed from 2.84×10^6 independent clones from the COLNNOT07 library. Starting RNA was made from colon tissue removed from a 60-year-old Caucasian male during a left hemicolectomy. The normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9228-9232), Swaroop et al. (Nucl. Acids Res. (1991) 19:1954) and Bonaldo et al. (Genome Res (1996) 6: 791-806), using a significantly longer (48 hour) reannealing hybridization period.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	Program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	Program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs*: Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least ten functions: fasta, tfastn, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs*: fasta E value= 1.06E-6 *Assembled ESTs*: fasta Identity = 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences*: fastx score=100 or greater
BLIMPS	Blocks Improved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	Pfam Hidden Markov Model-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	algorithm that searches for structural and sequence [illegible] in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score = 4.0 or greater
Phred	base-calling algorithm that examines automated [illegible] ces with [illegible] sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	Phils Revised Assembly Program including SWAT and crossMatch programs based on efficient implementation of Smith-Waterman algorithm, useful in searching [illegible] homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length= 56 or greater
Consed	graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	[illegible] matrix analysis program that scans protein [illegible] ces for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that match those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

TABLE 6

Nu SEQ	ide NO:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
			443531H1	1	253
			1406807F6	152	336
		43531	443531T6	847	355
			SBBA00451F1	396	856
			SBBA00676F1	546	865
			632860H1	13	253
		32860	784715R3	17	666
			509590H1	455	706
		70010	670010H1	1	263
			669971R1	1	633
			726498H1	13	263
		26498	726498R6	13	489
			866599R3	7	660
			795064H1	86	323
			4339458H1	4	284
		95064	937605R3	86	505
			2381151F6	592	1057
			1466346F6	857	1241
			924925H1	111	412
		24925	3268330H1	2	239
			759120R3	111	629
			1907958F6	1	478
		62390	023569F1	1122	470
			167282F1	1216	543
			1309211F1	911	1224

TABLE 6 (cont.)

Nu ie SEC NO:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
	1259405	1259405H1	46	277
		2472425H1	331	354
		774303R1	190	743
		1520779F1	418	1001
		1693833F6	914	1467
		831858T6.comp	1336	1742
		1527737T6.comp	1386	1829
143	1297384	1297384H1	402	641
		1269310F6	1	492
		1457367F1	792	1380
		415587R1	1358	1712
		SANA02967F1	1143	614
	1299627	1299627H1	1	250
		1359140F6	1004	1573
		1349224F1	1330	1731
		SBAA01431F1	46	397
		SBAA02909F1	868	262
		SBAA01156F1	901	1266
	1306026	1306026H1	1	223
		1464088R6	302	829
		SBAA02496F1	92	568
		SBAA04305F1	366	883
146	131 19	1316219H1	246	491
		2458603F6	1	402
		2504756T6	980	380
	1329031	1329031H1	1	264
		1329031T6	505	1
		1329031F6	1	523

TABLE 6 (cont.)

Nu... SEQ ... :	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
		1483050H1	722	931
		855049H1	1	267
		077017F1	1069	679
	1483050	1483050F6	722	1215
		1480024T6	2063	1315
		1483050T6	2068	1535
		759486R1	1762	2089
		1514160H1	1640	1838
		1866765T7	2383	2210
		782676R1	1652	1875
	1514160	008055X4	1090	1804
		008055X5	1316	1952
		1866765F6	2209	2391
		SAOA03127F1	2129	1703
		1603403H1	7	224
	1603403	372910F1	420	44
		733299R7	219	420
		1652303H1	4	256
		1671806H1	1	224
		1341743T1	2069	1900
		3803812H1	389	697
151	1652303	1878546F6	747	1344
		1428640F1	1081	1664
		2058609R6	1715	2098
		1331621F1	1780	2096
		1306331T1	1897	2098

TABLE 6 (cont.)

Nuc SEQ	de O:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
			1693358H1	41	125
			2498265H1	1	252
		1693358	1867125F6	205	373
			1693358T6	1094	416
			2245848R6	737	1103
			1707711H1	408	626
			1484609T1	2165	1855
			1707711F6	408	987
		1707711	1267959F1	1721	2182
			1484609F1	1855	2178
			SAJA00930F1	544	1132
			SAJA01300R1	1675	1212
			SAJA00999R1	1675	1142
			1738735H1	7	236
		1738735	SAJA00944R1	393	5
			SAJA00137F1	913	685
			SAJA03629F1	435	42
			1749147H1	1	276
		1749147	1749147F6	47	457
155			1749147T6	479	1
156		1817722	1817722H1	1	268
			2011085H1	344	545
			1831290H1	10	257
			3473958H1	70	242
			1972268F6	163	617
157		1831290	1301277F1	413	852
			1521574F1	1024	1602
			1561690T6	1729	1058
			891461R1	1261	1738

TABLE 6 (cont.)

Nucle[illegible] SEQ ID NO:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
[illegible]	1831477	1831477H1	59	337
		1582867H1	1	199
		1336769T1	1986	1639
		1933092H1	525	789
		1519909F1	841	1296
		1220946H1	1061	1318
		809556T1	1983	1687
		1217559T1	2002	1445
		1309225F1	1747	2001
1	1841607	1841607H1	13	192
		SBHA03588F1	13	172
160	1852391	1852391H1	98	367
		734140H1	1	225
		1852391F6	98	542
1	1854555	1854555H1	1	265
		2511711H1	37	58
		782453R1	223	712
		1854555F6	1	346
		1840675T6	1046	860
		2109736H1	938	1054
16[illegible]	1855755	1855755H1	17	224
		3040236H1	1	179
		1283207F1	306	816
		833763T1	1148	835
		1920926R6	854	1161
[illegible]	1861434	1861434H1	13	253
		1861434T6	872	261
		SARA01525F1	426	808
		SARA02548F1	587	889

TABLE 6 (cont.)

Nucleotide SEQ ID NO:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
	1872334	1872334H1	1	229
		1872334F6	1	424
		SBGA03684F1	358	425
	1877230	1877230H1	1405	1677
		2519841H1	1	251
		1877230T6	1903	1405
		1254693F1	335	716
		077020R1	682	1414
		1232336F1	906	1507
		1004952R6	1451	1904
		SARA01879F1	1545	1921
		SARA02654F1	1545	1923
166	1877885	1877885H1	68	323
		508020F1	499	51
		2751126R6	219	516
		SARA02571F1	407	499
167	1889269	1889269H1	757	1020
		1915551H1	1	191
		629493X12	481	865
		1441289F1	693	865
		1215274X34F1	1106	1631
		1818447F6	1307	1540
		1208463R1	1372	1493
168	1890243	1890243H1	9	268
		SARA01884F1	521	168
		SATA00046F1	1057	851
		SARA03294F1	1329	910
		SARA02790F1	1138	1535

TABLE 6 (cont.)

Nu[illegible]e SEQ [illegible]):	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
[illegible]	1900433	1900433H1	1	242
		SATA00396F1	409	124
		SATA02742F1	1	294
[illegible]	1909441	1909441H1	786	1048
		1398811F1	1	550
		3039939H1	607	876
		3324740H1	685	944
		1442131F6	787	1232
		2254056H1	1423	1522
		2199453T6	1955	1351
		1698531H1	1968	1796
[illegible]	1932226	1932226H1	294	510
		2320569H1	1	266
		1932226F6	294	685
		2469455T6	1475	1071
		2469455F6	1034	1492
		1907140F6	1158	1482
		SATA02592F1	857	518
[illegible]	1932647	1932647H1	17	246
		1492745T1	1582	1418
		1492745H1	1418	1599
		SASA02355F1	386	19
		SASA00117F1	250	569
		SASA00192F1	515	816
173	2124245	2124245H1	45	190
		1235393F1	495	895
		1402264F6	323	925
		1303990F1	682	1240
		1402264T6	1613	950

TABLE 6 (cont.)

Nucleotide SEQ ID NO:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
174	2132626	2132626H1	406	651
		1723432T6	1299	746
		2132626R6	406	904
		1736723T6	1292	857
		1504738F1	868	1320
	2280639	2280639H1	28	303
		1377560F6	261	777
	2292356	2292356H1	717	968
		4086827H1	1	275
		1754442F6	232	577
		3571126H1	497	808
		1601305F6	808	1464
177	2349310	2349310H1	1	236
		2349310T6	682	2
178	2373227	2373227H1	298	524
		3316444H1	801	1053
		302685R6	1141	1496
		SASA02181F1	577	1
		SASA01923F1	963	466
		SASA03516F1	1102	1249
	2457682	2457682H1	1	226
		2457682F6	1	554
180	2480426	2480426H1	1	213
		2480426F6	1	501

TABLE 6 (cont.)

Nuc[illegible]e SEQ [illegible]NO:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
[illegible]	2503743	2503743H1	6	222
		1853909H1	1	272
		1517619F1	172	830
		1467896F6	540	1112
		490031F1	1647	1068
		1208654R1	1382	1633
		880544R1	1450	1648
[illegible]	2537684	2537684H1	434	682
		2005493H1	1	194
		730969H1	307	547
		916487H1	723	989
		996135R1	997	1598
		1920738R6	1306	1692
		1957710F6	1472	1692
183	2593853	2593853H1	1	252
		807497H1	2	217
		914020R6	284	740
		889992R1	416	729
[illegible]	2622354	2622354H1	3	266
		2623992H1	1	246
		1556510F6	81	258
185	2641377	2641377H1	126	369
		4341415H2	10	345
		SBCA07049F3	126	599

TABLE 6 (cont.)

Nucleotide SEQ ID NO:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
[illegible]	2674857	2674857H1	139	393
		1872373H1	1	270
		470512R6	1486	1502
		1728547H1	1285	1508
		3013651F6	1423	1987
		SBCA01366F1	819	385
		SBCA00694F1	973	1198
[illegible]	2758485	2758485H1	20	267
		3097533H1	1	158
		1578959F6	291	771
188	2763296	2763296H1	63	301
		3486025F6	1	130
		SBDA07002F3	63	687
189	2779436	2779436H1	1	233
		2779436F6	1	577
		SBDA07009F3	1	608
19[illegible]	2808528	2808528H1	25	335
		2611513F6	2	489
		SBDA07021T3	1058	443
191	2809230	2809230H1	409	630
		2213849H1	1	133
		711706R6	396	691
		958323R1	407	800
		030732F1	1366	623
192	2816821	2816821H1	210	501
		3746964H1	1	307
		2816821F6	210	682
		948722T6	959	527

TABLE 6 (cont.)

Nucleotide SEQ ID NO:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
1	2817268	2817268H1	42	282
		3591308H1	13	264
		419522R1	179	808
		2073028F6	446	924
		1308781F6	869	1112
1	2923165	2923165H1	8	295
		2011630H1	18	238
		1457250F1	268	856
		754668R1	327	878
		1406510F6	558	901
19	2949822	2949822H1	1	280
		SBDA07078F3	1	606
1	2992192	2992192H1	25	321
		2534324H2	1	240
		2815255T6	690	219
		1551107T6	893	471
		1551107R6	471	690
1	2992458	2992458H1	48	362
		2618951H1	1	247
		1479252F1	163	610
		1879054H1	563	840
		1879054F6	563	1096
		2215240H1	951	1202
		1535968T1	1729	1173

TABLE 6 (cont.)

Nucl[illegible]e SEQ [illegible]O:	Clone ID	Fragment of SEQ ID NO	Starting Nucleotide of Fragment	Ending Nucleotide of Fragment
19[illegible]	3044710	3044710H1	652	952
		3741773H1	1	283
		859906X42C1	94	192
		1534347F1	90	268
		1421122F1	830	1392
		1303865F1	1033	1487
		1704452F6	1432	1934
		1251642F1	2006	1544
		1781694R6	1894	2017
199	3120415	3120415H1	72	363
		1360123T1	523	141
		1375015H1	380	526